safety

MDPI

Article

Deep Learning for Detection of Proper Utilization and Adequacy of Personal Protective Equipment in Manufacturing Teaching Laboratories

Adinda Sekar Ludwika and Achmad Pratama Rifai *

Department of Mechanical and Industrial Engineering, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia; adindasekarludwika@mail.ugm.ac.id
* Correspondence: achmad.p.rifai@ugm.ac.id

Abstract: Occupational sectors are perennially challenged by the potential for workplace accidents, particularly in roles involving tools and machinery. A notable cause of such accidents is the inadequate use of Personal Protective Equipment (PPE), essential in preventing injuries and illnesses. This risk is not confined to workplaces alone but extends to educational settings with practical activities, like manufacturing teaching laboratories in universities. Current methods for monitoring and ensuring proper PPE usage especially in the laboratories are limited, lacking in real-time and accurate detection capabilities. This study addresses this gap by developing a visual-based, deep learning system specifically tailored for assessing PPE usage in manufacturing teaching laboratories. The method of choice for object detection in this study is You Only Look Once (YOLO) algorithms, encompassing YOLOv4, YOLOv5, and YOLOv6. YOLO processes images in a single pass through its architecture, in which its efficiency allows for real-time detection. The novel contribution of this study lies in its computer vision models, adept at not only detecting compliance but also assessing adequacy of PPE usage. The result indicates that the proposed computer vision models achieve high accuracy for detection of PPE usage compliance and adequacy with a mAP value of 0.757 and an F1-score of 0.744, obtained with the YOLOv5 model. The implementation of a deep learning system for PPE compliance in manufacturing teaching laboratories could markedly improve safety, preventing accidents and injuries through real-time compliance monitoring. Its effectiveness and adaptability could set a precedent for safety protocols in various educational settings, fostering a wider culture of safety and compliance.

Keywords: PPE detection; deep learning; computer vision; YOLO; laboratory

Citation: Ludwika, A.S.; Rifai, A.P. Deep Learning for Detection of Proper Utilization and Adequacy of Personal Protective Equipment in Manufacturing Teaching Laboratories. *Safety* **2024**, *10*, 26. https://doi.org/10.3390/safety10010026

Academic Editor: Raphael Grzebieta

Received: 12 January 2024
Revised: 15 February 2024
Accepted: 1 March 2024
Published: 7 March 2024

1. Introduction

Awareness regarding workplace safety in developing countries remains relatively low. Especially in Indonesia, this is evident through the high number of insurance claims with the Social Security Administrative Agency of Indonesia (BPJS Ketenagakerjaan) for Employment Accidents, totaling 234,370 claims due to workplace accidents in 2022 [1]. The causes of these accidents are manifold, often attributed to inadequate planning, production organization, unsafe workplace conditions, and human factors. These human factors may stem from psychological origins or reflect societal, cultural, and organizational training issues [2]. The lack of compliance with Standard Operating Procedures (SOPs) regarding the use of Personal Protective Equipment (PPE) plays a role in these accidents, influencing the overall safety level. Studies suggest that lower frequency of PPE use corresponds to higher chances of workplace accidents [3].

PPE serves as a safeguarding tool used by workers to protect themselves from occupational hazards. Utilizing PPE stands as a critical strategy in preventing injuries and occupational illnesses stemming from workplace hazard exposure, which can result from direct contact with chemicals, radiation, physical, electrical, and mechanical risks, and

other workplace dangers. Among the most significant methods to safeguard worker health and safety in the workplace against potential risks or hazards is the proper use of PPE [4].

The National Institute for Occupational Safety and Health (NIOSH), a United States federal agency tasked with the role of conducting research and formulating recommendations to prevent work-related injuries and illnesses, integrates the use of PPE within the Hierarchy of Controls, a methodology to determine the most effective actions for controlling exposure to hazards. This hierarchy comprises five levels of control actions, ranked by effectiveness: elimination, substitution, engineering controls, administrative controls, and PPE. When other control methods fall short in reducing hazardous exposure to safe levels, organizations must provide PPE. Moreover, PPE is frequently used in situations where hazards are not adequately controlled, making it the last line of defense in creating a safe environment when other controls prove insufficient.

It is essential to acknowledge that accidents can occur not only in the workplace but also in educational settings during practical learning activities. Generally, student learning activities mostly occur within classrooms; however, they also take place in laboratories. Laboratory-based learning involves practical experiences directly related to the objects within the lab. During laboratory activities, students must adhere to SOPs, or the procedures outlined for that specific laboratory. Notably, the use of PPE is crucial in manufacturing teaching laboratories in universities. This is because these types of spaces house machinery and equipment that could pose risks to student safety.

The potential hazards within the laboratory arise from the interaction between humans and machinery. Manufacturing teaching laboratories house various machines such as CNC machines, lathes, shapers, milling machines, drills, grinders, electric welding machines, saws, and wire cutting machines. These machines have fast-moving, automated components. Additionally, tools like hammers, chisels, and saws have shapes and usage patterns that pose risks to users. Accidents in the laboratory encompass scenarios such as eye exposure to metal shavings, skin injuries due to metal dust contact, slipping on spilled lubricants, eye injuries from welding exposure, clothing being pulled into machines, finger injuries from chisels, saws, and grinders, chuck dislodging incidents, falling objects hitting the head, hand injuries from lathes, falling laboratory tools injuring feet, hand injuries during machine operation, and hands getting trapped in machinery.

Given the occurrence of such accidents, the proper use of PPE becomes mandatory for students' compliance and safety. However, issues related to the use of PPE by students persist, including inadequate usage and varying levels of compliance. Incomplete PPE usage might involve not wearing all the required pieces. Improper usage, on the other hand, encompasses using PPE in ways that deviate from the prescribed guidelines. The necessary PPE for manufacturing teaching laboratories include hard hats, lab coats, safety shoes, masks, safety glasses, earmuffs, and gloves. Activities such as welding and grinding necessitate the simultaneous use of these seven types of PPE. However, compliance with these PPE types extends beyond concurrent usage; students in the manufacturing teaching laboratories should adhere to wearing all seven PPE types during their activities.

To ensure the completeness and accuracy of PPE usage, this study develops detection models using artificial intelligence, especially deep learning. Convolutional Neural Networks (CNNs) are employed to classify whether students are wearing PPE appropriately and completely during activities in the manufacturing teaching laboratories. CNNs possess the ability to recognize and detect objects within digital images, largely due to enhanced computational power, large datasets, and improved training techniques [5]. The challenge, however, lies in maintaining high detection accuracy. Addressing this challenge, the You Only Look Once (YOLO) algorithm proves effective in object detection [6].

Based on this, this research focuses on identifying the level of completeness and accuracy in the usage of PPE in the manufacturing teaching laboratories in universities by developing a CNN detection model using the YOLO algorithms. The main objective is to implement a detection system that optimally identifies PPE objects within a detection frame. This visual-based object detection process relies on the presence of PPE objects in the image,

and the number of detected PPE objects in a single frame is unaffected by the number of individuals. As such, the detection system is developed to identify the PPE object that can be implemented in images with single or multiple individuals within a single frame. The goal is for this system implementation to accurately and automatically detect PPE objects used by students and laboratory users, ensuring that PPE usage conforms appropriately and dutifully to its intended purpose.

2. Literature Review

PPE serves the vital purpose of safeguarding its wearer's body from occupational hazards and preventing accidents. However, several factors, including low awareness of PPE usage, discomfort, fatigue, and negligence, contribute to insufficient PPE utilization and incorrect handling among workers [7]. Research conducted by [8] concerning accurate PPE usage detection at construction sites has highlighted the potential of computer–vision-based methods for automatically detecting PPE completeness. These methods offer non-invasive, cost-effective perception on-site, as they typically identify all workers and PPE components before verifying if a worker is using PPE based on their relationship with the involved equipment.

The need to detect PPE arises in order to enhance the completeness and accuracy of its usage, thereby reducing workplace accidents [9]. Consequently, there exists a significant practical requirement to utilize technology that can assist practitioners in improving or ensuring PPE completeness. While efforts to incorporate electronic circuits into PPE have been made [10], technologies allowing visual and noncontact completeness adherence to safety regulations are more prevalent and practical to implement. To accommodate this requirement, deep learning becomes a prominent approach for PPE detection.

In research by [6], a CNN was used to concurrently detect workers' use of hard hats and vests. Overall, CNN-based methods directly process images of workers and classify the status of PPE usage, including both the level of completeness and accuracy, through an end-to-end inference process. A research conducted research on PPE detection with the aim of reducing workplace accidents in the construction industry [11]. Their study employed a CNN to detect PPE usage by workers and classify various types of PPE, such as determining if each worker wears a hard hat. Beyond hard hats, some studies have extended PPE detection to various tools, with simultaneous detection processes. Ref. [12] detected multiple types of PPE, including hard hats and vests, using a CNN for a comprehensive safety assessment. Moreover, Ref. [6] used a CNN to simultaneously detect workers' use of hard hats and vests. Overall, CNN-based methods directly process images of workers, classifying both the level of completeness and accuracy of PPE usage.

Earlier research on PPE completeness and accuracy detection using deep learning extensively analyzed and reviewed various deep learning algorithms employed in developing systems aimed at identifying PPE usage. The selection of these algorithms is based on the methods or techniques used in the developed systems to identify the presence of PPE objects. In a study conducted by [13] which focused on PPE detection at construction sites, the YOLO detection method was used to identify hard-hat-wearing personnel. The detection of PPE objects was carried out using YOLOv3 and YOLOv4. In the PPE detection process, the proposed model offered practical detection performance in terms of speed and accuracy. This method holds significant potential for automated inspection of PPE components. Based on testing results, it was able to achieve detection efficiency of over 25 FPS and a mAP value of 97%, which can be utilized to ascertain whether construction personnel adhere to safety regulations and meet real-time, high-accuracy requirements. This demonstrates that YOLOv3 possesses high accuracy and detection speed, while YOLOv4 outperforms YOLOv3, particularly in terms of detecting small objects with improved speed and accuracy.

The YOLO algorithm exists in multiple versions, each with its own performance characteristics. In a study conducted by [14], the focus was on PPE detection at construction sites, specifically considering the use of YOLOv5. Their study aimed to detect PPE usage

among construction workers across six PPE categories: shoes, jackets, vests, gloves, glasses, and hard hats. The performance of the proposed YOLOv5s model variant was compared to other algorithms through three indicators: precision, recall, and F1 score. Comparative algorithms included YOLOv4, Faster-RCNN MobileNetV3, and Faster-RCNN Resnet50. The results of a five-fold cross-validation technique revealed that YOLOv5s exhibited the most effective performance in terms of precision and recall indicators. The enhanced YOLOv5 model yielded the highest precision and recall values compared to benchmark models.

In research by [9] conducted in the manufacturing industry, several deep learning architectures were considered: MobileNetV2, VGG19, Dense-Net, Squeeze-Net, InceptionV3, and ResNet. All algorithm models were pretrained on the ImageNet dataset and implemented using the PyTorch framework. The study was conducted within the manufacturing sector to classify various types of PPE used by workers, including hard hats, gloves, and protective glasses. The obtained performance indicated that MobileNetV2, Dense-Net, and ResNet were the top-performing classifiers. These three models achieved comparable performance, with MobileNetV2 offering the added advantage of being the most computationally efficient. For hard hat classification, both MobileNetV2 and ResNet showed superior performance, with an average accuracy of 95%.

PPE detection within the manufacturing sector was also explored by [15], focusing on the detection of hard hats. The algorithms used in their study were YOLOv4 and YOLOv4-Tiny. YOLOv4-Tiny is a lightweight version of the complete YOLOv4, explicitly designed to reduce object detection time. A dataset of 7112 labeled images was used to discern images containing workers wearing hard hats or not. Results showed that, for the YOLOv4 configuration, the hard hat class achieved an AP50 score of 96.09%. In YOLOv4-Tiny, the AP50 score reached 86.53% for hard hats. The trade-off between accuracy and complexity of these two networks is evident. While YOLOv4 achieved the highest object recognition level, the YOLOv4-Tiny network demonstrated the best latency in object detection tasks.

Furthermore, a study conducted by [16] employed a ReID model for accuracy in detecting and inferring the usage of PPE by each identified worker. For ReID, a novel loss function called similarity loss was designed to encourage the deep learning model to learn more discriminative human features. By combining ReID and PPE classification results, a workflow was developed to record incidents of workers not wearing the required PPE. With a real construction site dataset, the proposed method improved worker ReID and PPE classification accuracy by 4% and 13%, respectively, facilitating site video analysis and safety compliance inspection among workers. For the ReID component, three algorithm models were utilized: ResNet50, OSNet, and OSNet + BDB. The results revealed that ResNet50 exhibited the highest accuracy performance with a precision of 97.91%.

Based on the research by [8] regarding accurate PPE usage detection at construction sites to prevent potential hazards, the employed algorithm utilized the OpenPose model for worker pose estimation. To expedite the detection process, an optimized method adopted the MobileNet network as a feature extractor. This network employed depth-wise separable convolution filters to separate depth and spatial information, enhancing computational efficiency. The study took into consideration worker poses and body postures for detection. The inclusion of pose consideration aimed to estimate key human points, localize body regions and heads based on these key points, and utilize image classification for PPE detection. The proposed worker detection method achieved a precision of 99.61% and a recall of 98.04% in worker detection, indicating that 0.39% of workers were falsely detected, and 1.96% of workers were missed in the images.

Apart from construction sites, the detection of PPE objects is also performed in offshore drilling operations. A study conducted by [17] aimed to propose a framework for PPE detection. The proposed framework aimed to enhance the accuracy, reliability, and performance of PPE detection compared to existing methods. The detected PPE objects were safety hard hats and workwear for offshore drilling. The framework was built using YOLOV4 and evaluated using accuracy, recall, false alarm rate (FAR), missed alarm rate (MAR), and detection time as evaluation metrics. The proposed method was then

compared with other methods from the literature, and experimental results indicated that the proposed framework outperformed other methods for PPE detection on offshore drilling platforms. The accuracy values for safety hard hat detection was 87.8%, while for workwear, the accuracy value was 93.1%. This framework could detect workers not wearing safety hard hats or workwear in a timely manner and generate alarm messages. However, a limitation of the study was that the detection system did not accurately identify PPE when the heads or torsos of some workers were partially obscured by pipes during operations. This could lead to inaccurate detection results and reduce the overall accuracy of the framework. Additionally, extreme weather conditions such as fog and heavy rain at sea could result in blurry images, making it difficult to accurately locate and identify workers, thereby reducing PPE detection accuracy.

The application of PPE object detection is also carried out in public spaces, as these spaces pose exposure risks that can impact health. A study conducted by [18] detected PPE objects such as face shields, face masks, and gloves. The algorithm used was YOLOv4. The study utilized a total of 8000 iterations and resulted in a mAP value of 79% for all detected PPE classes and a loss value of 2.97. Overall, the YOLOv4 algorithm proved to be a fast and accurate model suitable for object detection monitoring purposes.

The development of CNN models in relation to this topic has been carried out in various work settings. The most frequently studied work location is construction sites. Some studies have also developed models based on datasets from nuclear power plants, industrial workplaces, factories, and public spaces. Previous studies that conducted PPE detection, such as those research that focused on construction site PPE detection [6,8,13,19,20]. Meanwhile, research by [9], Refs. [14,15] conducted in industrial settings, and [17] conducted research in the offshore drilling environment. The common objective of these previous studies was to detect PPE to prevent workplace accidents. Similarly, other studies, like that of [18] aimed to detect PPE in public areas to prevent human exposure to hazards, thus necessitating the use of PPE for prevention. In contrast, the focus of this study is on PPE detection in manufacturing teaching laboratories in universities to prevent workplace accidents.

In high-risk work environments such as industrial factories, construction sites, and nuclear power plants, rigorous monitoring of PPE usage is commonplace. This is largely attributed to well-defined SOPs and structured activity schedules. In contrast, manufacturing teaching laboratories in universities often exhibit less stringent oversight of PPE compliance. Ref. [21] surveyed the compliance of PPE usage, in which the results indicate that respondents from academics are significantly less compliant with wearing a lab coat (66%) and eye protection (61%) than respondents from government labs (73% and 76%, respectively) and from industry (87% and 83%, respectively). Ref. [22] further collected data of incidents in industrial and university labs, in which a majority of the incidents (65%) were taking place in universities. This discrepancy stems from the lack of formalized SOPs specifically for PPE and the variable nature of individual visitation schedules. Given these unique challenges, there is a compelling need for an advanced PPE detection system. The identification of PPE compliance and adequacy can be achieved through a detection system that employs deep learning technology with a CNN, given the unique challenges posed by this context. Although several studies have been proposed to develop PPE detection in various industrial sites, to the best of our knowledge this is the first study dedicated to the development of a CNN-based PPE detection model specifically for laboratory environments. Hence, the purpose of this study is to develop a CNN model for detecting compliance and adequacy of PPE usage among visitors in manufacturing teaching laboratories in universities.

The distinction of this study from previous related research lies in its execution within manufacturing teaching laboratories in universities, whereas prior studies were conducted in construction sites, industrial settings, power plants, and nuclear facilities. The differentiating factor from earlier research, despite a shared focus on PPE detection, lies in the diverse environmental backgrounds of each research location. The varying background

context of detection locations leads to differing detection outcomes even when utilizing the same detection methodology. Additionally, there are differences in the type of PPE used in industrial and university laboratories. This study encompasses seven types of PPEs, going beyond detecting the presence of PPE objects in images to providing precision levels of usage for each individual piece of PPE. The PPE classes investigated include hard hats, lab coats, shoes, masks, protective eyewear, and gloves. Furthermore, this study advances the development of detection models that are capable of not only identifying PPE objects in use but also evaluating their correct utilization. This factor is particularly critical in university settings, as students and laboratory users typically receive less training on PPE usage compared to workers in industrial and construction sectors. Consequently, this leads to a higher incidence of incorrect use of PPE within academic environments.

This study develops PPE detection models by harnessing the YOLO algorithms, noted for their efficacy and precision in related fields. Our research evaluates three versions: YOLOv4, YOLOv5, and YOLOv6, each chosen for its distinctive strengths. The inclusion of these variants is crucial to comprehensively assess their performance in PPE detection, especially within the consistent setting of manufacturing teaching laboratories in universities. Despite the uniform environment, variations in the algorithms' design can lead to different model outputs, underscoring the need to explore the effectiveness of each YOLO version.

Empirical evidence suggests varied strengths among these algorithms: YOLOv6 often surpasses its predecessors in general performance, largely owing to its more advanced development. However, this does not diminish the notable precision and high mean Average Precision (mAP) score of YOLOv4 [23], nor the exceptional detection speed of YOLOv5 [24]. Conversely, YOLOv6, while more accurate overall [20], has shown limitations in close-up object detection and stability compared to YOLOv5 [20]. To address these conflicting findings and further our understanding of these algorithms' capabilities in PPE detection, our study conducts a detailed comparative analysis of YOLOv4, YOLOv5, and YOLOv6.

3. Detection Method

3.1. YOLOv4

YOLO is a deep learning technique that excels in rapid, accurate object detection and classification, adapting well to detection tasks [25]. Unlike two-stage algorithms that separate object detection into distinct phases, YOLO encapsulates the process in a single pass through the neural network, analyzing an image in one stride to predict object locations and classifications. This streamlined approach allows YOLO to offer real-time performance benefits, making it highly efficient for applications requiring immediate detection results. YOLO's architecture mimics human neural networks, enabling it to recognize and detect objects by leveraging previously learned data, thus exhibiting remarkable proficiency when presented with new, unseen images. This efficiency and adaptability render YOLO a preferred choice over other object detection algorithms, especially in scenarios where speed and accuracy are important.

The YOLOv4 algorithm, developed by [26], represents an advancement over its predecessor, YOLOv3, and operates as a one-stage detector, demonstrating a remarkable capability for rapid object detection [26]. YOLOv4 consists of three essential components: the backbone block, neck block, and head block. The backbone block serves the purpose of feature extraction from images. Subsequently, the neck block enhances feature effectiveness by introducing additional layers between the backbone and head/neck. In YOLOv4, the head block is responsible for identifying and classifying objects within each bounding box, achieved by applying anchor boxes to the feature map and generating a final output vector containing class probabilities, object scores, and bounding box information [27]. Figure 1 presents an illustration of the YOLOv4 architecture.

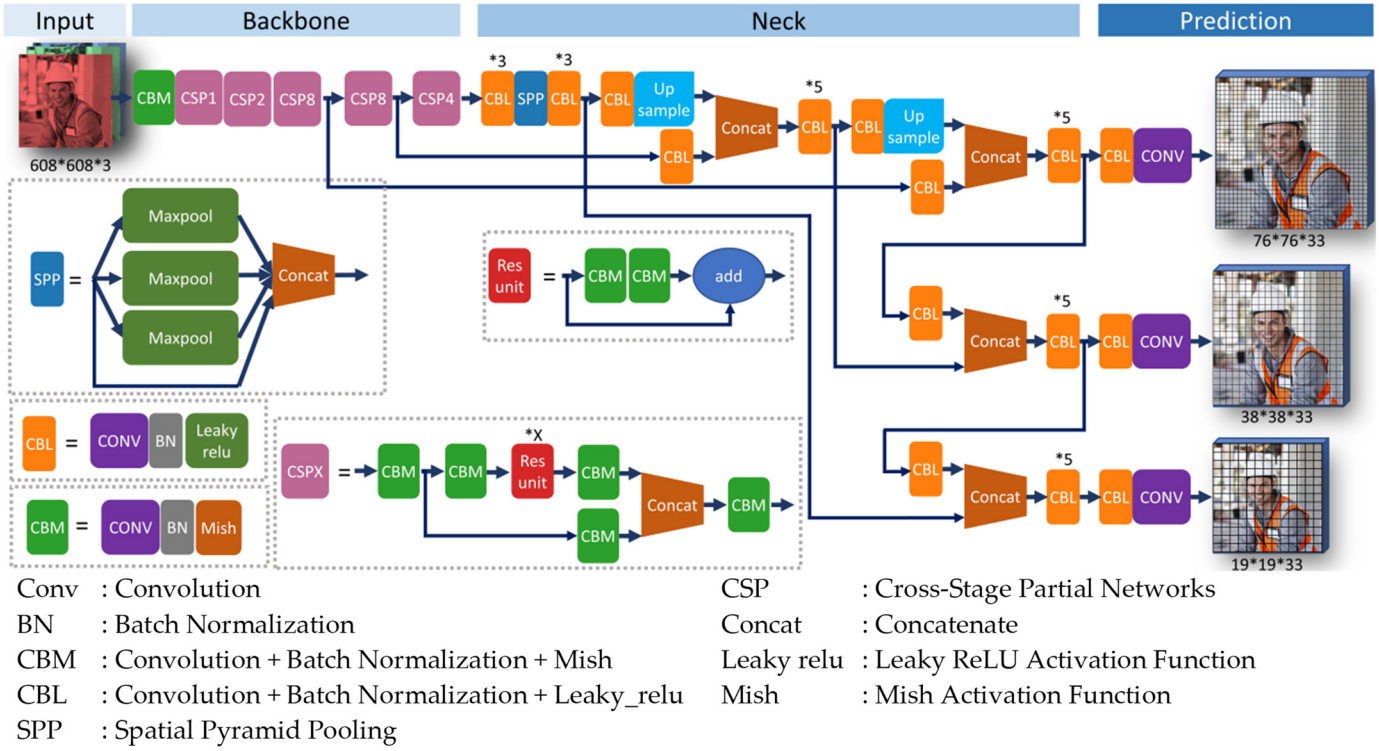


Figure 1. Architecture of YOLOv4 (Adapted from [28]).

The YOLOv4 architecture initiates its processing pipeline with an input layer that accommodates an image of 608 by 608 pixels, encoded in three color channels (RGB). This input is then subjected to a series of convolutional operations within the custom CBM blocks, where each convolutional layer is followed by batch normalization and the Mish activation function, thereby allowing for initial feature extraction and nonlinear activation.

Subsequent to the initial feature extraction, the architecture employs CSP blocks, specifically CSP1, CSP2, and CSP8. These blocks ingeniously partition the feature maps, process each segment independently, and subsequently merge the results. This strategic design is aimed at reducing computational overhead while enhancing the efficiency of the feature extraction process.

Interspersed with these CSP blocks are sequences of CBL modules, each comprising convolutional layers, batch normalization, and Leaky ReLU activation functions. These modules are replicated multiple times within the network (denoted by the symbols *3 or *X), indicating their repeated application for the purpose of feature refinement.

The SPP layer, situated after the CSP8 block, introduces a critical enhancement to the architecture. By executing max-pooling operations at varying scales, the SPP layer aggregates contextual information and ensures a comprehensive receptive field, which is vital for capturing features at different scales and resolutions.

Upon completion of the backbone's feature extraction phase, the architecture's 'Neck' serves as a sophisticated feature fusion mechanism. It employs upsampling and concatenation operations to amalgamate feature maps from disparate layers. This fusion process allows the network to consolidate information across different resolutions, thereby enabling more accurate object detection across varying scales.

In the concluding 'Prediction' phase, the YOLOv4 model employs convolutional layers at three distinct scales, specifically 76 × 76, 38 × 38, and 19 × 19. These scales correspond to the network's ability to capture small, medium, and large object details, respectively. At each scale, the convolutional layers predict bounding boxes, objectness scores, and class probabilities. The output of the network is a set of bounding boxes and associated class

predictions, which are superimposed onto the original input image, effectively highlighting the detected objects.

The incorporation of mosaic data augmentation, along with optimizations in the backbone, network training, activation functions, and loss functions, elevates YOLOv4 as a robust algorithm for object detection. Inspired by its previous ability in object detection, this study employs YOLOv4 for the object detection algorithms for PPE detection. In this study, the YOLOv4 model is constructed through the utilization of transfer learning and employs a loss function comprising both box loss and classification loss. Originally, YOLOv4 had undergone training on the Common Objects in Context (COCO) dataset, encompassing over 200,000 labeled images across 80 distinct classes. Its utilization in detecting PPE objects within manufacturing teaching laboratories in universities is attributed to YOLOv4's ability to strike a balance between speed and accuracy in object detection [23].

3.2. YOLOv5

The YOLOv5 object detection algorithm is a one-stage, anchor-based object detection method. Its architecture comprises three key components: the backbone, neck, and head. With its comprehensive design, the YOLOv5 model is proficient in learning the characteristics and image features from the provided dataset, allowing it to perform object detection based on the features acquired during the training process [20]. The architecture of YOLOv5 is illustrated in Figure 2.

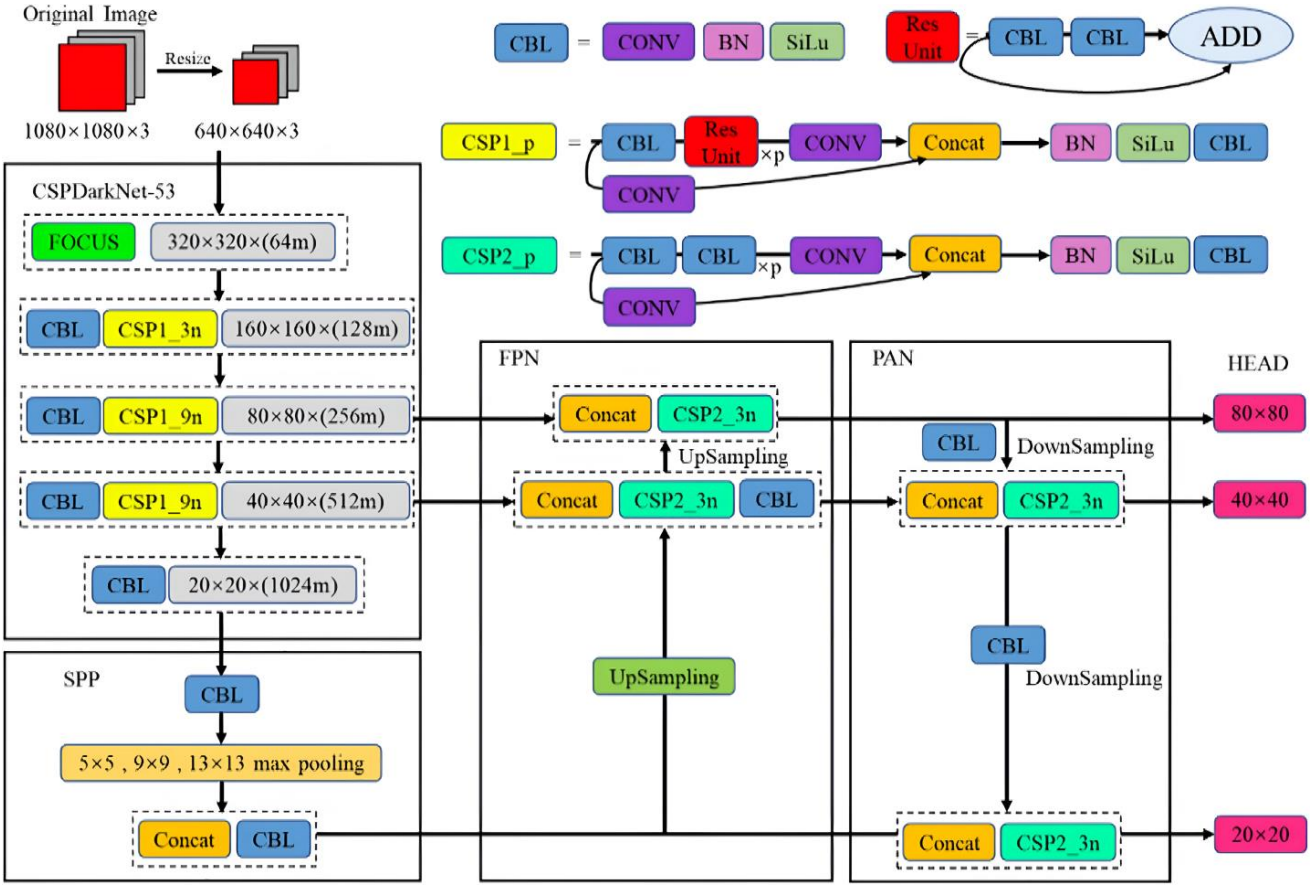


Figure 2. Architecture of YOLOv5 (Adapted from [29]).

The first stage of processing is undertaken via the CSPDarkNet-53 backbone, which comprises a series of CBL layers, interspersed with CSP blocks. The CSP blocks, denoted by CSP1 with varying multiplicative factors (3n and 9n), serve to manage computational efficiency by splitting the feature map into two parts, processing each independently, and then recombining them. This structure is repeated at multiple scales, progressively

halving the spatial dimensions while doubling the depth of the feature maps, as seen in the transition from 320 × 320 to 20 × 20 pixels.

The SPP layer follows, which aggregates the feature maps using max-pooling operations at 5 × 5, 9 × 9, and 13 × 13 scales. This operation captures spatial hierarchies and robust features at varying scales, leading to a concatenated feature map that is fed into additional CBL layers.

The architecture then transitions into the Neck, composed of a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN). The FPN utilizes top-down and lateral connections to merge feature maps from different levels of the backbone, enhancing semantic information at each scale. The PAN, conversely, facilitates the bottom-up pathway, enhancing the feature hierarchy by aggregating lower-level spatial details through a series of upsampling and concatenation operations. This structure ensures that finer-grained features are preserved and enhanced as they flow through the network.

Finally, the detection heads, referred to as the HEAD in the diagram, process the aggregated feature maps to produce object detections at three distinct scales: 80 × 80, 40 × 40, and 20 × 20. These scales correspond to different grid sizes where the network predicts bounding boxes, object classes, and confidence scores. The multi-scale detection capability allows YOLOv5 to effectively detect objects of various sizes within the image.

In this research, YOLOv5 serves as one of the selected algorithms. Here, the transfer learning approach is also used to develop the YOLOv5 model for PPE detection. Similar to YOLOv4, YOLOv5 also originally underwent training on the Common Objects in Context (COCO) dataset, an extensive collection of data used for object recognition, segmentation, and labeling [30]. The selection of YOLOv5 for detecting PPE objects in manufacturing teaching laboratories in universities is justified by its excellent sensitivity in object detection [29].

3.3. YOLOv6

The YOLOv6 CNN is an object detection algorithm that operates in a single stage, identifying objects in images without the need for preliminary regional proposal network (RPN) processing. This leads to enhanced detection speed, accuracy, and model parameter reduction. Figure 3 presents an illustration of the YOLOv6 architecture.

The model architecture is methodically organized into a series of stages, each building upon the preceding one to incrementally refine and enhance the feature representation of the input image. At the inception of the model, an input image with a resolution of 1280 × 1280 pixels is passed through a stem layer, which is a convolutional module designed to initiate the feature extraction process. This layer prepares the image for deeper processing within the network.

Progressing into the model, the architecture is delineated into multiple stages, each composed of various convolutional modules and CSP layers. Convolutional modules, characterized by specific kernel sizes, padding, and stride values, are responsible for detecting patterns and features at different spatial hierarchies of the input image. BN and the SiLU (Sigmoid Linear Unit) activation function are consistently applied after convolutional operations to stabilize the learning process and introduce nonlinearity.

The architecture further incorporates specialized bottleneck modules, namely DarknetBottleneck and SPPFBottleneck. These modules are designed to further condense and filter the feature maps, focusing the model's attention on the most salient features. The DarknetBottleneck employs a residual structure, allowing gradients to flow more effectively during training, while the SPPFBottleneck leverages spatial pyramid pooling to capture contextual information at various scales.

As the processed features flow through the network's backbone, they are advanced into the Neck, which is composed of additional convolutional modules and CSP layers. The culminating section of the model, the Head, is tasked with the critical role of generating predictions. It processes the aggregated feature maps and applies convolutional layers to

predict object bounding boxes, class probabilities, and confidence scores across various scales, resulting in precise localization and identification of objects within the image.

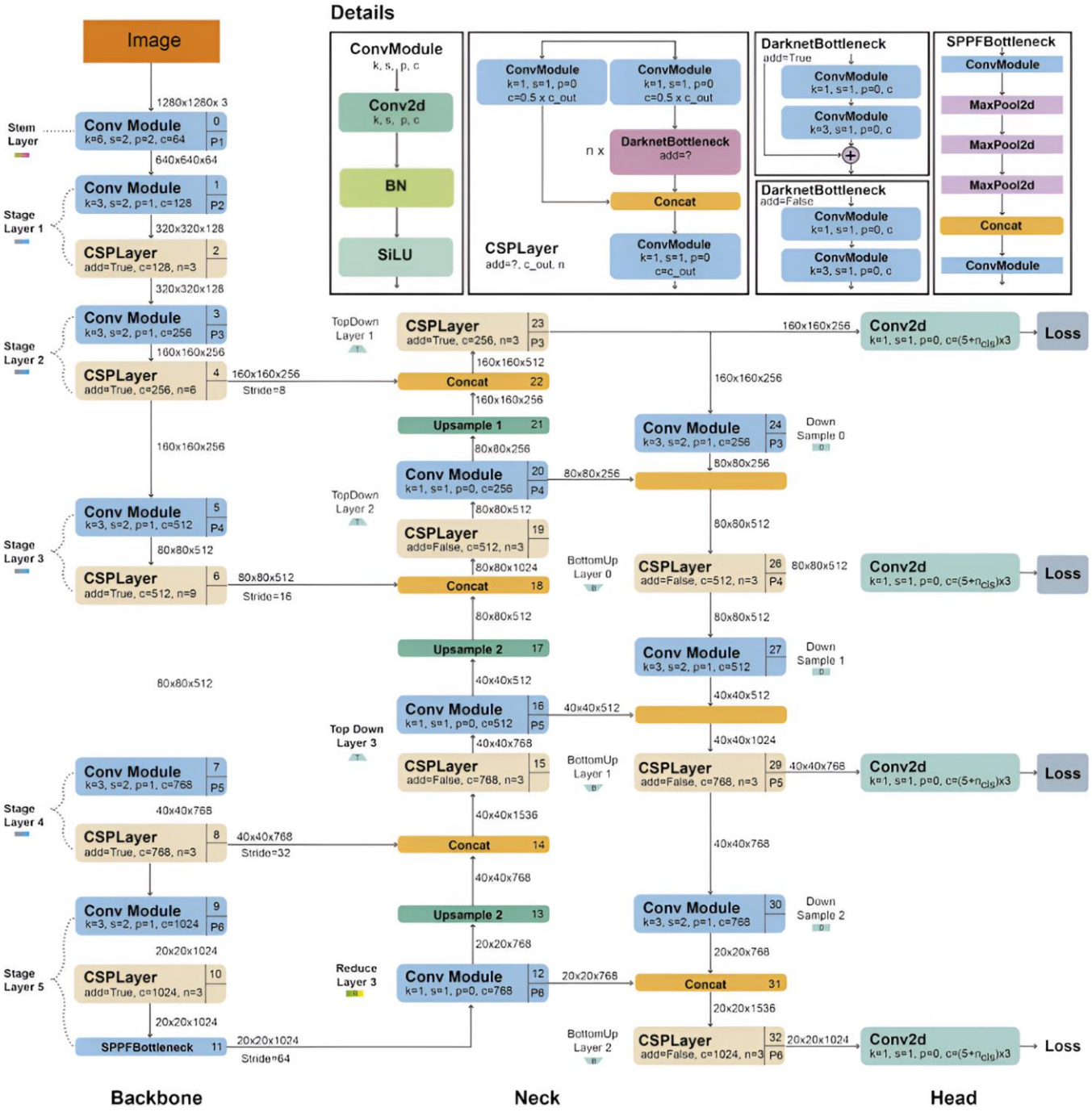


Figure 3. Architecture of YOLOv6 (Adapted from [31]).

Each stage of the architecture is intricately linked to the next, ensuring that the feature representation becomes increasingly refined. The model outputs a set of predictions, which are compared against the ground truth during training using loss functions, guiding the network to minimize errors and improve detection performance.

In this research, YOLOv6 is selected as one of algorithms for detecting PPE objects within a university laboratory setting. The selection of YOLOv6 is attributed to its notable advantages, including improved processing speed and object detection accuracy when

compared to its predecessors, ranging from YOLOv1 to YOLOv5 [25]. Similar to the previous method, the transfer learning approach is chosen for training the YOLOv6 model.

3.4. Evaluation Metrics

The test results are analyzed and interpreted to assess the performance of all CNN models, addressing the research questions and objectives defined in advance. The analysis involves comparing the outcomes of the three algorithms using predefined evaluation metrics. These evaluation metrics are utilized to assess and analyze the model's performance and include precision, recall, F1 score, accuracy, Average Precision (AP), and mean Average Precision (mAP). Additionally, for the training results, an analysis is conducted based on the loss curve to gauge the accuracy of the object detection, taking into consideration the components of loss, which comprise box loss (loss related to bounding boxes), object loss (loss related to object detection), and class loss (loss related to class prediction). The detailed criteria for performance evaluation of the three models both in training and in the testing process are shown in Table 1.

Table 1. Evaluation criteria.

No	Parameter	Indicator
1	Training result	Accuracy, precision, recall, F1 score, and mAP obtained via the models during the training process Evaluation of the possibility of overfitting or underfitting
2	Testing result	Accuracy, precision, recall, F1 score, and mAP from overall classes on the test datasets Evaluation of confusion matrix and AP of each class

Indicator	Formula	Description
Accuracy	$\text{Accuracy} = \frac{TP}{TP+FP+FP+FN}$	Accuracy is a measure of the correctness of a model's predictions. This metric provides a straightforward indication of performance across all classes.
Precision	$\text{Precision} = \frac{TP}{TP+FP}$	Precission tells how many objects are correctly predicted to be positive. This metric determines whether or not the model is reliable.
Recall	$\text{Recall} = \frac{TP}{TP+FN}$	Recall is a metric to measuring how well the model identifies the number of objects detected correctly against the number of ground truth objects.
F1 Score	$\text{F1 Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$	F1 assesses the trade-off between precision and recall. When the F1 value is large, it shows that the precision and recall are large and vice versa. The best score for the F1-score was 1.0 and the worst score was 0.
mean Average Precision (MAP)	$\text{mAP} = \frac{\sum_{i=1}^{N} AP_i}{N}$	mAP is the average of AP values for all classes in a dataset. This provides an overview of the model's performance over the various classes.

4. Experiment Setting

4.1. Dataset

The focus of this study involves testing seven categories of PPE, namely hard hats, lab coats, safety shoes, masks, safety glasses, earmuffs, and gloves. These seven types of PPE classes are among the most used PPE in manufacturing teaching laboratories in universities. The use of these seven categories of PPE has been deemed sufficient to protect the user's body in manufacturing teaching laboratories since the activities of the machines used are not in the form of heavy machinery activities commonly found in the industrial setting that require more comprehensive protection. For each PPE category, multiple variations or color options are considered. For instance, within the hard hat category, different colors are explored, while in the case of gloves, safety glasses, earmuffs, and masks, variations extend to both shape and color. The images in Figures 4–10 showcase the seven PPE categories that will be the subject of examination.

The dataset consists of 4150 images captured with various variations in distance, angles, lighting conditions, and backgrounds. These images encompass both single individuals and multiple individuals. In the case of multiple individuals, the image captures typically involve groups ranging from 3 to 5 individuals. Regarding the angles of image capture employed in data collection, there are two distinct angles used. The first angle is positioned at a camera height that can encompass an area containing 3–5 individuals within a single frame. Meanwhile, the second angle is aligned at the same height as the individuals under study.

Figure 4. Hard hats.

Figure 5. Lab coat.

Figure 6. Safety shoes.

Figure 7. Masks.

Figure 8. Safety glasses.

Figure 9. Earmuffs.

Figure 10. Gloves.

4.2. Data Preprocessing

4.2.1. Data Annotation

After the data are collected, each image undergoes data labeling, also known as annotation, which is the process of assigning labels to each class within all images in the dataset. Data annotation serves the purpose of making the training process a reliable and effective source of learning for the computer before subjecting the model to testing by detecting elements in the testing dataset. Additionally, data annotation forms the foundation for evaluating the accuracy of predictions against the labels assigned to each image in the training dataset during testing.

The data annotation is conducted manually using the Roboflow software tool. From the whole dataset, there are 16,119 annotated objects. Gloves are the class with the greatest number of labels, while shoes have the fewest. Since this study also detects the correctness of PPE utilization, there are two classes for each PPE, both representing correct and incorrect utilization. As such, there are a total of 14 classification labels for the PPE detection problem. Figure 11 presents an example of data annotation in an image.

In Figure 11, there are several observed instances of improper PPE usage. The gloves are worn incorrectly, with the rough part intended for the palm positioned on the back of the hand instead. The hard hat is not securely fastened to the head, raising the risk of it easily coming off. Furthermore, the individual in the image is wearing footwear that does not meet safety shoe standards, thus the shoes are classified as 'incorrect'. Table 2 presents the number of annotated objects in the datasets for all classes.

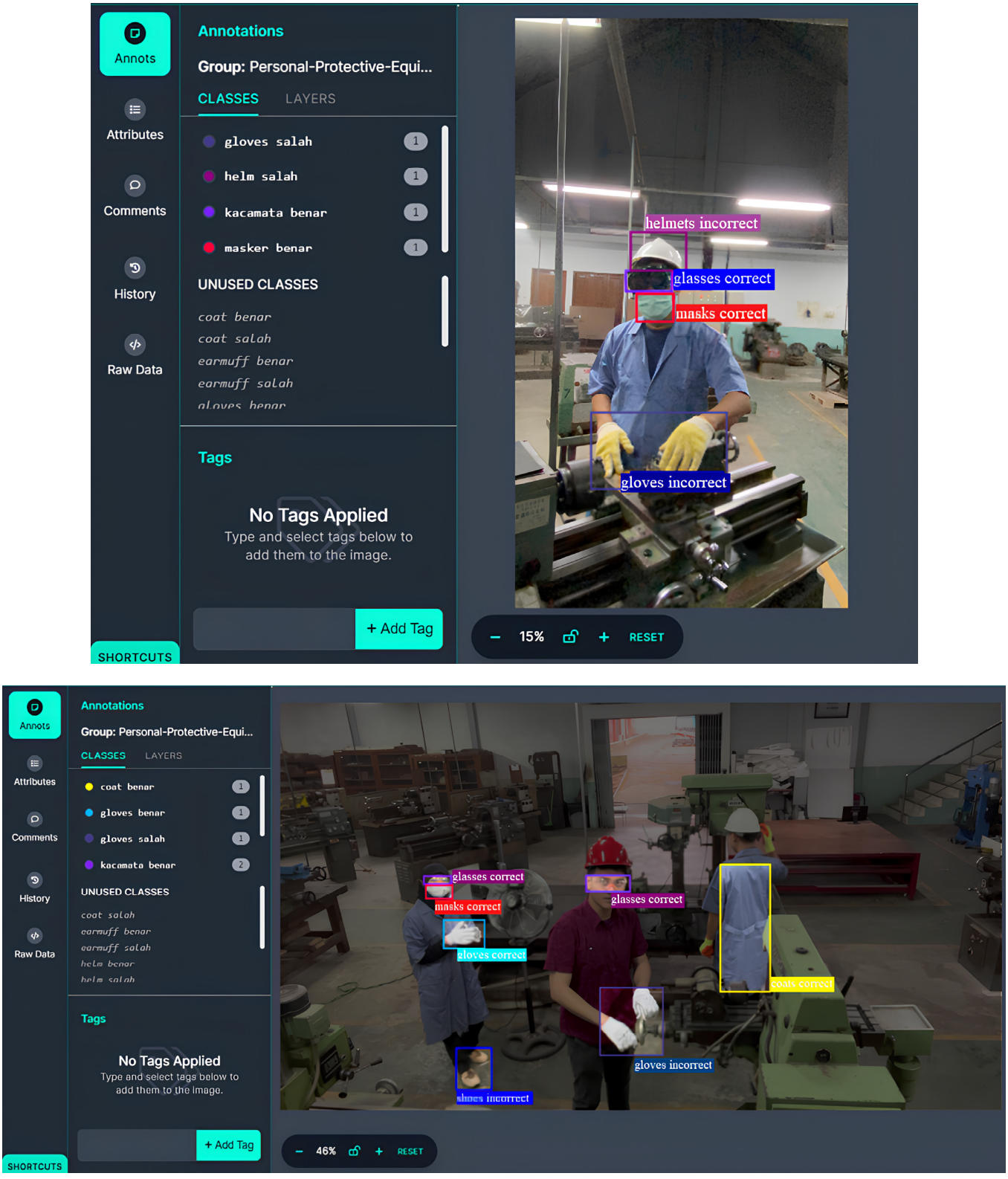


Figure 11. Example of data annotation.

Table 2. Number of annotated objects in each class.

PPE Object	Class	Number of Annotated Objects
Gloves	Correct utilization	2803
	Incorrect utilization	601
Coat	Correct utilization	1834
	Incorrect utilization	1308
Glasses	Correct utilization	1503
	Incorrect utilization	1008
Earmuff	Correct utilization	205
	Incorrect utilization	168
Safety shoes	Correct utilization	485
	Incorrect utilization	911
Mask	Correct utilization	2166
	Incorrect utilization	237
Hard hats	Correct utilization	2570
	Incorrect utilization	320

After annotation, the dataset is divided into two categories based on their roles in constructing the CNN model: training dataset (70%) and testing dataset (30%). The training dataset is employed for the model's learning or training process. The validation dataset is used to validate the model and evaluate its performance. It is crucial for the validation dataset to be distinct from the training dataset.

4.2.2. Data Augmentation

Data augmentation is the process of expanding the dataset by modifying images to enhance its diversity. This procedure is exclusively applied to the training dataset. Data augmentation is carried out using the Roboflow software, employing various augmentation techniques available within it. In this study, augmentation is performed by adjusting parameters such as saturation, blur, flip, noise, brightness, and exposure in the images. Table 3 presents the configuration and settings for each augmentation function. After the augmentation process, the number of images in the training dataset increase to 8709 images.

Table 3. The configuration and settings for each augmentation function.

Function	Configuration
Flip	Horizontal and vertical
Rotate	Clockwise and counter-clockwise by 90 degrees
Brightness	−10% until +10%
Exposure	30% darker or brighter
Blur	Up to 1 pixel
Noise	Up to 2% of total pixels

Besides expanding the number of images, each filter serves its unique purpose, aimed at introducing greater diversity to the images. Adjusting image saturation can significantly enhance the model's performance when dealing with varying lighting conditions in different background images. The application of exposure filters is aimed at training the object detection model to be more resilient to lighting variations that may occur due to changes in lighting and camera settings. Noise filters represent one method for training the object detection model to better handle variations and disturbances within images. These filters are utilized to introduce granular details to the images. Small pixel alterations within these filters can prevent overfitting and lead to accurate predictions. Noise can simulate real-world conditions where images may experience distortion or interference, posing challenges in object detection.

Blur filters are employed to bolster the reliability of the object detection model by imparting an out-of-focus effect to the images. This helps the model become more robust to

camera focus, teaching it to recognize objects even in blurry or hazy conditions. Brightness filters are used to adjust the brightness levels in images, either enhancing or reducing them. The use of these filters ensures that the model can effectively identify objects in both well-lit and low-light image conditions. This diversification ensures that during the detection process the model will yield more robust results. This is achieved by augmenting the dataset with a variety of filters that alter the images in terms of color, brightness, clarity, and sharpness, resulting in increased image variability.

4.3. Hyperparameter Setting

To ensure a fair comparison of each detection model, the hyperparameters of learning rate *lr*, batch size *B*, loss function, and optimizer are set constant throughout the experiments. Besides that, we further investigate the effect of the number of epochs on the performance of the models, and its influence on overfitting and underfitting possibilities during model training. The loss function choices are cross-entropy loss for YOLOv4 and YOLOv5, and VariFocal Loss for YOLOv6. The descriptions of the hyperparameter are as follows:

- Learning rates control how much model weights change each time a training iteration is performed. The learning rate function is to help algorithms achieve faster and better convergence during model training to ensure that the algorithm does not jump too far or too slowly in finding the minimum or optimum point of the loss function;
- Batch refers to a group of data samples processed simultaneously in an iteration. Batch enables better computational efficiency and better handling of fluctuations in data;
- Epoch is a stage in which algorithms perform one complete iteration of model training with all existing datasets. One epoch means that the model has seen and learned part of the training data in a batch (sample group) of all available training data;
- Image size is the size dimension of an input image used to detect objects in an image;
- An optimizer is a function that adjusts attributes such as weights and learning rate. Thus, it helps to reduce overall loss and improve object detection capabilities.

The hyperparameter settings used for the training process are shown in Table 4. The detection models are developed and evaluated in Google Collaboratory written in Python with the PyTorch and TensorFlow libraries.

Table 4. Hyperparameter values.

Hyperparameter	Value
number of epochs	50, 75, 100
batch size *B*	32
Image Size	416 × 416
learning rate *lr*	0.001
optimizer	SGD

5. Experiment and Result

After the datasets for training and validation, both with single or multiple individuals in the images, had been formed, the next step was to use the training dataset to build the models and then observe the architecture performance based on the validation dataset. Three models are generated based on YOLOv4, YOLOv5, and YOLOv6. Afterward, the performance of the models is evaluated based on obtained precision, recall, and mAP. Further, the model performance in detecting each PPE objects is assessed by evaluating the confusion matrix.

5.1. Results of YOLOv4

Three models are developed using YOLOv4, with the number of epochs set to 50, 75, and 100. Training processes are executed using the hyperparameter values described in Section 4. The computation times are longer with the increase in the number of epochs. The training process for the YOLOv4 detection models took 285 min for the 50 epochs model,

337 min for the 75 epochs model, and 450 min for the 100 epochs model. However, the increase in the number of epochs also allows the model to learn more on the pattern of the datasets, as indicated by the increase in evaluation metrics; albeit, the increase is not significant. Table 5 presents the value of the evaluation metrics both for all classes and in each of the testing datasets for the YOLOv4 models.

Table 5. Results for the YOLOv4 models.

Class		50 Epochs Model				75 Epochs Model				100 Epochs Model			
		Precision	Recall	mAP 0.5	F1 Score	Precision	Recall	mAP 0.5	F1 Score	Precision	Recall	mAP 0.5	F1 Score
All Classes		0.614	0.607	0.621	0.61	0.627	0.616	0.647	0.621	0.643	0.624	0.657	0.633
Coats	Correct utilization	0.645	0.632	0.622	0.638	0.742	0.766	0.773	0.754	0.752	0.757	0.785	0.754
	Incorrect utilization	0.629	0.614	0.605	0.621	0.812	0.746	0.798	0.778	0.824	0.724	0.811	0.771
Earmuffs	Correct utilization	0.314	0.315	0.323	0.314	0.468	0.487	0.386	0.477	0.497	0.486	0.394	0.491
	Incorrect utilization	0.421	0.345	0.413	0.379	0.711	0.49	0.592	0.58	0.732	0.535	0.612	0.618
Gloves	Correct utilization	0.656	0.618	0.642	0.636	0.781	0.674	0.75	0.724	0.796	0.683	0.692	0.735
	Incorrect utilization	0.524	0.498	0.502	0.511	0.636	0.414	0.494	0.502	0.654	0.487	0.519	0.558
Hard hats	Correct utilization	0.75	0.786	0.797	0.768	0.75	0.852	0.831	0.798	0.782	0.835	0.864	0.808
	Incorrect utilization	0.607	0.659	0.679	0.632	0.617	0.673	0.696	0.644	0.638	0.692	0.705	0.664
Glasses	Correct utilization	0.658	0.603	0.618	0.629	0.716	0.675	0.691	0.695	0.721	0.705	0.712	0.713
	Incorrect utilization	0.667	0.616	0.634	0.64	0.719	0.654	0.683	0.685	0.739	0.648	0.696	0.691
Masks	Correct utilization	0.814	0.857	0.842	0.835	0.829	0.873	0.894	0.85	0.837	0.881	0.873	0.858
	Incorrect utilization	0.789	0.746	0.725	0.767	0.716	0.751	0.799	0.733	0.874	0.766	0.783	0.816
Safety shoes	Correct utilization	0.356	0.303	0.315	0.327	0.563	0.525	0.514	0.543	0.575	0.547	0.521	0.561
	Incorrect utilization	0.572	0.467	0.458	0.514	0.705	0.569	0.581	0.63	0.718	0.578	0.584	0.64

The results indicate that the YOLOv4 models undergo underfitting during the training process since the evaluation metric values are relatively low. The results also indicate that there is a significant gap in the obtained evaluation metrics values for each class. This occurrence is influenced by significant differences in the number of objects within each class. For example, the average precision in the masks correct class is significantly higher than the earmuff correct class since the masks correct class has a significantly higher number of annotated objects (2166 annotation), while the earmuff correct class has 205 annotations. Similar results were also observed for recall and F1 score values, in which the class with the higher number of annotated objects has relatively higher metric values.

The results presented in the table indicate that among the three types of epochs used in the YOLOv4 model, the highest F1 score is achieved for the epoch 100 model. The selection of the best model is based on the F1 score because it encapsulates both precision and recall values. Additionally, the mAP, which serves as a quality measure for an object detection system by considering precision at various threshold levels for different objects, attains its highest value for the epoch 100 model. This suggests that among the three epoch models, the 100 epochs model exhibits the best performance in detecting PPE objects among the YOLOv4 models.

5.2. Results for YOLOv5

The training processes for the YOLOv5 models took significantly lower computation time than the YOLOv4 models, albeit with the same number of epochs. The training process for the YOLOv5 detection models took 78 min, 120 min, and 155 min for the YOLOv5 models with the number of epochs set to 50, 75, and 100 epochs, respectively. Figure 12 presents the loss curves for the YOLOv5 models obtained during the training process.

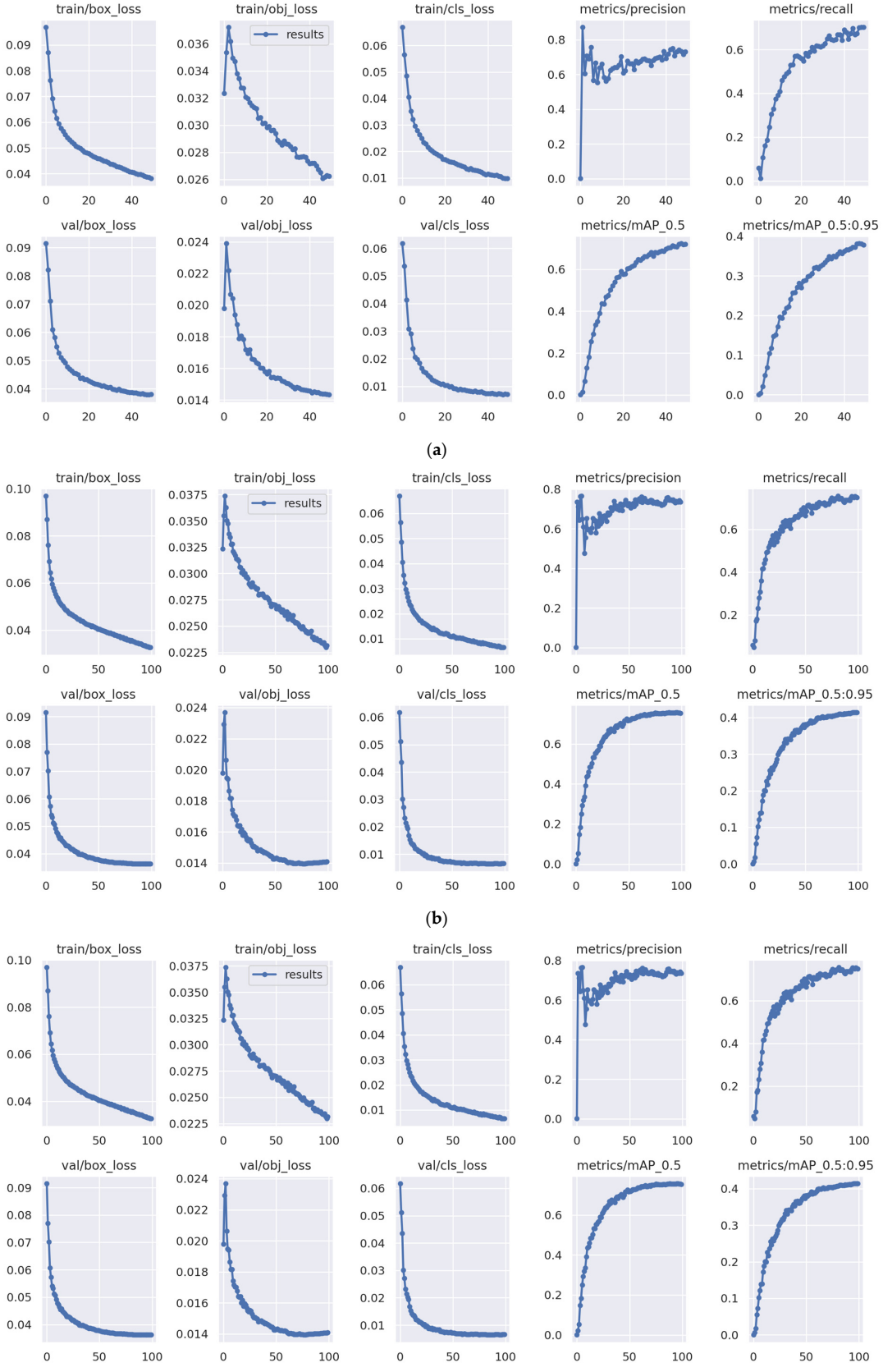


Figure 12. Loss curves for YOLOv5 models. (**a**) YOLOv5 50 epochs model, (**b**) YOLOv5 75 epochs model, and (**c**) YOLOv5 100 epochs model.

The curves in Figure 12 illustrate the changes in the loss function values during the training and validation phases of the model. The loss function serves as a metric for assessing how accurately the model makes predictions. The primary goal during model training is to optimize the loss function so that the model can make increasingly accurate predictions over time. Ideally, the loss values on the curve will decrease as training progresses, indicating that the model is improving its understanding of patterns in the data.

Referring to Figure 12, we can observe that the curves for box loss, object loss, and class loss on both training and validation exhibit a decreasing trend in all models. This signifies that the models are getting better at predicting box locations, recognizing objects within the image grid more precisely, and classifying objects into their correct categories. This also indicates progress in the models' comprehension of the structure and characteristics of objects in the images. In the 75 and 100 epochs models, there is a consistent positive trend in the training and validation data, without significant fluctuations in the validation data, thus indicating that there was no overfitting. However, in the 50 epochs model, it can be observed that the decreasing trend is terminated before reaching convergence both in training and validation loss, which implies that the training process of the model might be underfitting. This finding is also confirmed by the curves of the metric values, in which the 50 epochs model generally obtained lower precision, recall, and mAP at the end of the training process than the 75 and 100 epochs models. Further analysis is performed by evaluating the F1–confidence, precision–confidence, and precision–recall curves which are presented in Figures 13–15.

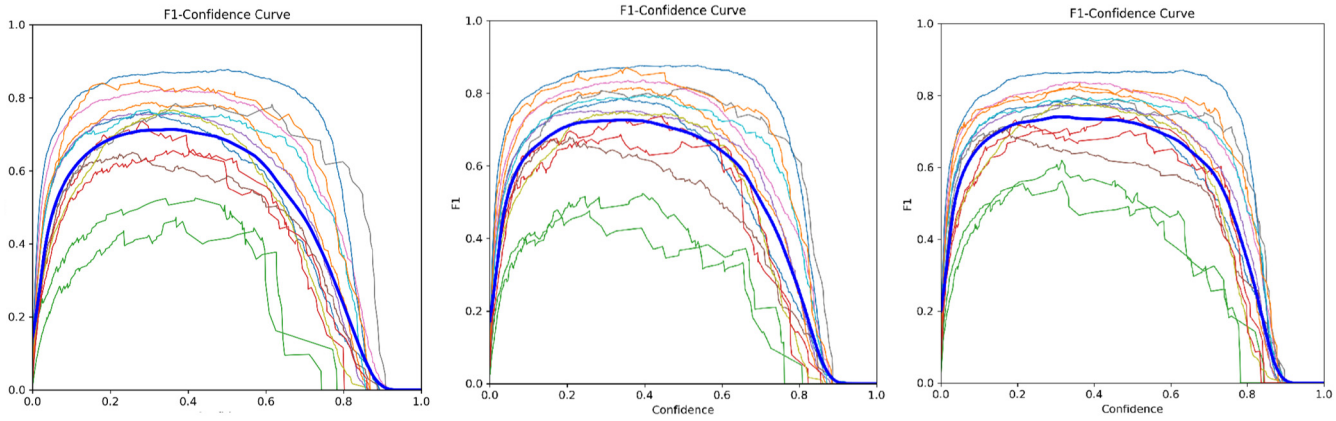


Figure 13. F1–confidence curve of YOLOv5 models. (**a**) 50 epochs model, (**b**) 75 epochs model, and (**c**) 100 epochs model.

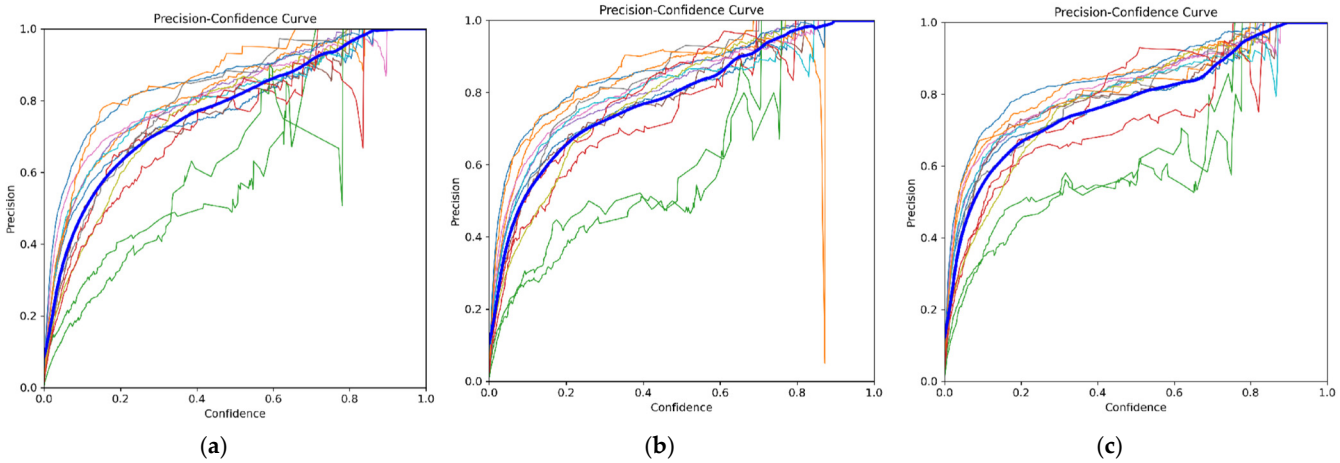


Figure 14. Precision–confidence curve of YOLOv5 models. (**a**) 50 epochs model, (**b**) 75 epochs model, and (**c**) 100 epochs model.

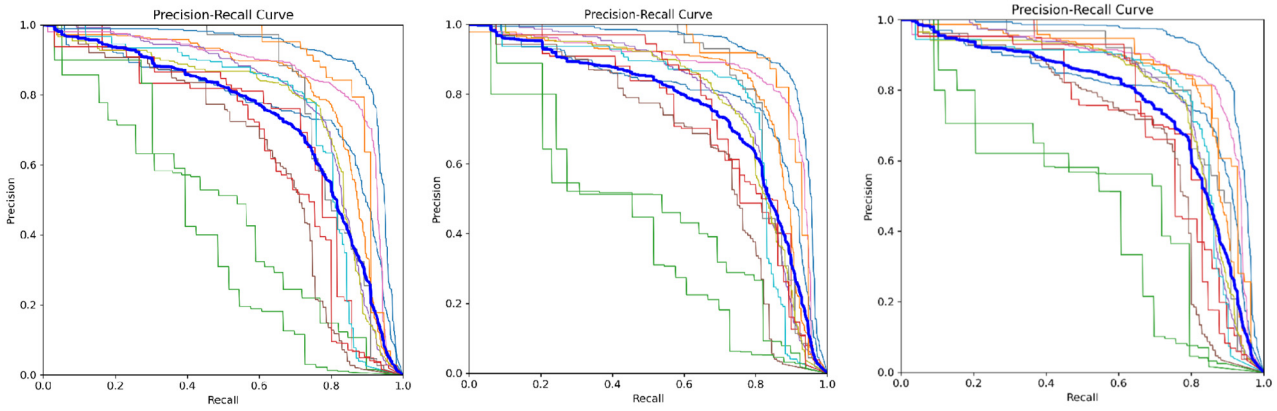


Figure 15. Precision–recall curve of YOLOv5 models. (**a**) 50 epochs model, (**b**) 75 epochs model, and (**c**) 100 epochs model.

In Figures 13–15, the color of each curve represents each PPE classes, providing a visualization of their respective values. Meanwhile, the dark blue curve symbolizes the data for all PPE classes. Figure 13 illustrates that a higher F1 score indicates superior model performance in PPE detection across a broad range. Figure 14 suggests that a higher position of the curve correlates with increased model accuracy in PPE prediction. Meanwhile, Figure 15 demonstrates the precision–recall relationship; the closer a curve is to the top edge, the better the balance between precision and recall, indicating optimal model performance.

The F1–confidence curve provides an insight into how well the model can make accurate predictions for the positive class while considering its ability to identify most of the positive objects present. For example, for the 50 epochs model, it is noted that the peak F1 score is 0.71 for all classes at a confidence level of 0.354. This implies that at this specific point on the curve, the classification model performs quite well, which means that the model is proficient at classifying positive and negative instances, achieving a balanced precision and recall. The peak F1 score obtained with the 75 epochs model is 0.73 at the same confidence level. While the 100 epochs model yields a peak F1 score of 0.74 at a confidence level of 0.31. Hence, indicating that, in general, both models performed better than the 50 epochs model.

In Figure 14, the precision–confidence curve for the YOLOv5 50 epochs model yields a result in which all classes have a precision value of 1.00 at a confidence level of 0.924. This outcome indicates that, based on the precision–confidence curve, all classes exhibit a precision of 1.00 (100%) at a confidence level of 0.924. This means that at this point on the curve, the classification model demonstrates exceptionally high precision performance, meaning that every time the model makes a positive prediction at a confidence level of 0.924, it is always correct (no false positives). Meanwhile, for the 75 epoch models, 100% precision is obtained at a confidence level of 0.898, while in the 100 epochs model 100% precision is obtained at a confidence level of 0.904. As such, in this aspect, the 75 epochs model is favorable since it can obtain high precision at a lower confidence level. This underscores that the model can return completely correct predictions with a relatively lower value of confidence.

The results for the precision–recall curves also strengthen the claim that the 75 and 100 epochs models generally are more favorable since they score higher in mAP, with values of 0.749 and 0.757, respectively, at a threshold of 0.5, as compared to the 50 epochs model with a mAP value of 0.724 at a threshold of 0.5. In addition, an analysis is also performed for evaluating the ability of the models to predict each class, which is depicted in Table 6.

Table 6. Results for YOLOv5 models.

Class		50 Epochs Model				75 Epochs Model				100 Epochs Model			
		Precision	Recall	mAP 0.5	F1 Score	Precision	Recall	mAP 0.5	F1 Score	Precision	Recall	mAP 0.5	F1 Score
All Classes		0.735	0.699	0.724	0.717	0.748	0.711	0.749	0.729	0.739	0.750	0.757	0.744
Coats	Correct utilization	0.730	0.764	0.767	0.747	0.760	0.808	0.793	0.783	0.734	0.831	0.795	0.779
	Incorrect utilization	0.793	0.762	0.811	0.777	0.838	0.789	0.824	0.813	0.833	0.804	0.832	0.818
Earmuffs	Correct utilization	0.487	0.564	0.461	0.523	0.455	0.538	0.480	0.493	0.543	0.640	0.524	0.588
	Incorrect utilization	0.758	0.639	0.692	0.693	0.687	0.628	0.721	0.656	0.656	0.694	0.723	0.674
Gloves	Correct utilization	0.778	0.732	0.779	0.754	0.780	0.724	0.784	0.751	0.774	0.769	0.797	0.771
	Incorrect utilization	0.713	0.534	0.622	0.611	0.746	0.541	0.660	0.627	0.781	0.553	0.681	0.648
Hard hats	Correct utilization	0.782	0.858	0.853	0.818	0.799	0.870	0.861	0.833	0.802	0.871	0.855	0.835
	Incorrect utilization	0.796	0.727	0.798	0.760	0.809	0.773	0.830	0.791	0.766	0.800	0.824	0.783
Glasses	Correct utilization	0.763	0.768	0.755	0.765	0.766	0.732	0.775	0.749	0.771	0.771	0.785	0.771
	Incorrect utilization	0.784	0.732	0.743	0.757	0.789	0.780	0.766	0.784	0.787	0.797	0.781	0.792
Masks	Correct utilization	0.847	0.895	0.910	0.870	0.844	0.899	0.919	0.871	0.829	0.896	0.921	0.861
	Incorrect utilization	0.834	0.806	0.884	0.820	0.904	0.839	0.897	0.870	0.781	0.857	0.866	0.817
Safety shoes	Correct utilization	0.535	0.394	0.418	0.454	0.508	0.364	0.411	0.424	0.567	0.517	0.466	0.541
	Incorrect utilization	0.685	0.615	0.647	0.648	0.783	0.668	0.765	0.721	0.721	0.692	0.750	0.706

The results indicated in Figures 13–15 show that there are some classes that have significantly lower metric values than all other classes with respect to prediction (highlighted with a green line), which are also confirmed by the result of validation shown in Table 6. Those classes are safety shoes correct utilization and earmuffs correct utilization. The same finding is observed in all YOLOv5 models (and also YOLOv4 models). This unfavorable performance is attributed to the limited number of annotated objects in both classes which hinders the training process of the model in learning the pattern of these two classes.

Moving on, the analysis is detailed by observing the confusion matrix presented in Figure 16. The safety shoes correct class obtained the lowest prediction accuracy at 0.48, 0.52, and 0.58 for the 50, 75, and 100 epochs models, respectively. In this class, the models fail to detect the correct safety shoes object, thus detecting this object as background. The same phenomenon also happened in the prediction of the earmuffs correct class. However, the 100 epochs model provides relatively better prediction with higher accuracy for these classes than the other two YOLOv5 models. It is also observed that all YOLOv5 models can differentiate the correct and incorrect utilization of PPE objects, indicated by a low value of wrong prediction via mistakenly detecting correct utilization as incorrect utilization and vice versa. The error is mostly attributed to the failure of detecting the object and assigning it as the background. Based on this finding, it can be safely claimed that the 100 epochs model is the best performed model using the YOLOv5 algorithm.

5.3. Results for YOLOv6

Similar to the other YOLO algorithms, three models are developed using YOLOv6, with the number of epochs set to 50, 75, and 100 epochs. The training time for the YOLOv6 models was 118 min, 137 min, and 190 min for the 50, 75, and 100 epochs models, respectively. Figures 17–19 present the loss curves both on training and validation datasets during the training process for YOLOv6 models.

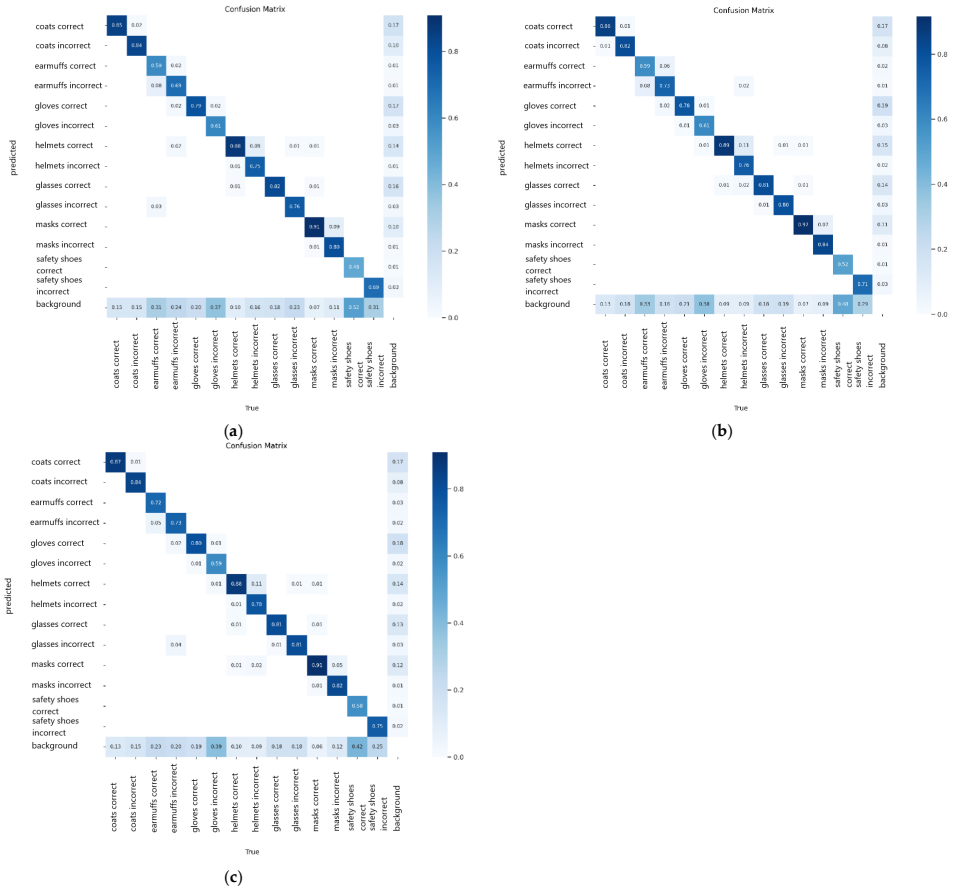


Figure 16. Confusion matrix of YOLOv5 models. (**a**) 50 epochs model, (**b**) 75 epochs model, and (**c**) 100 epochs model.

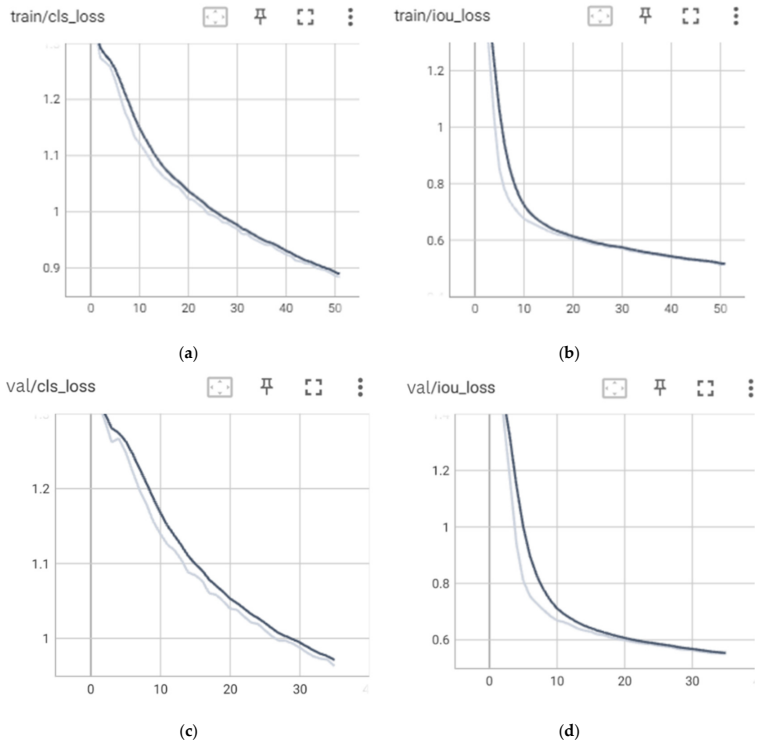


Figure 17. Loss curves of YOLOv6 50 epochs model. (**a**) Classification loss on training data, (**b**) IoU loss on training data, (**c**) classification loss on validation data, and (**d**) IoU loss on validation data.

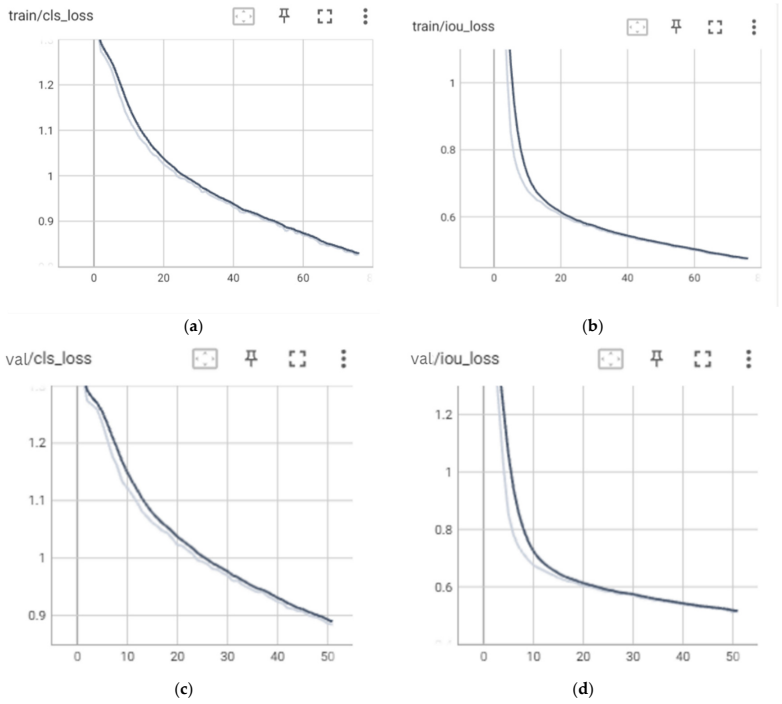


Figure 18. Loss curves of YOLOv6 75 epochs model. (**a**) Classification loss on training data, (**b**) IoU loss on training data, (**c**) classification loss on validation data, and (**d**) IoU loss on validation data.

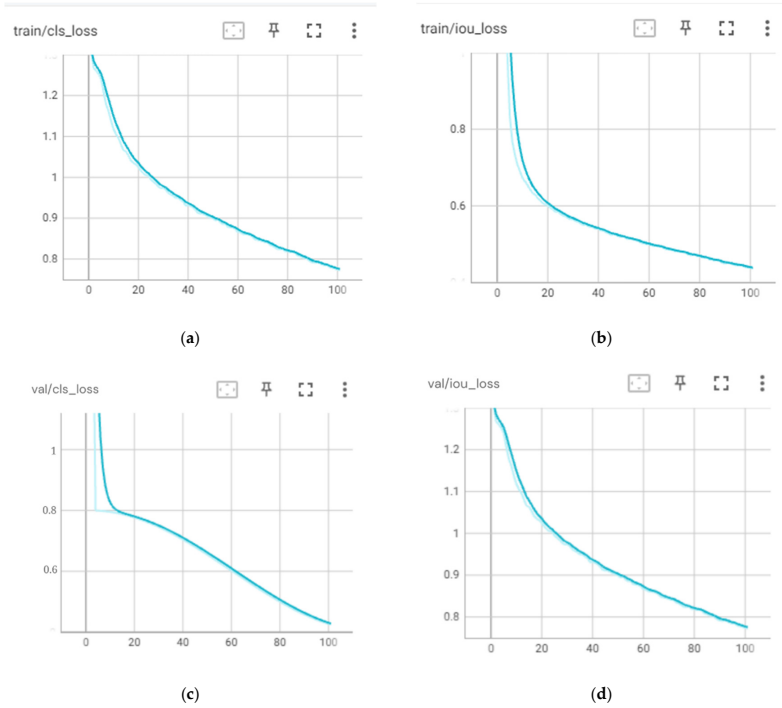


Figure 19. Loss curves of YOLOv6 100 epochs model. (**a**) Classification loss on training data, (**b**) IoU loss on training data, (**c**) classification loss on validation data, and (**d**) IoU loss on validation data.

Based on the figures, it can be observed that there is no indication of overfitting as the decreased trend of loss curves for both training and validation are consistent. However, there is an indication that the models encounter underfitting since the training stops before the loss curves reach convergence. As such, the YOLOv6 models might obtain lower loss if the number of epochs is increased and the training continues. Moving on, the performances of the YOLOv6 models in detecting the proper utilization and adequacy of each PPE class

for the testing datasets are evaluated. Table 7 presents the detailed results of the models, while Figure 20 illustrates the confusion matrix for the three models of YOLOv6.

Table 7. Results for YOLOv6 models.

Class		50 Epochs Model				75 Epochs Model				100 Epochs Model			
		Precision	Recall	mAP 0.5	F1 Score	Precision	Recall	mAP 0.5	F1 Score	Precision	Recall	mAP 0.5	F1 Score
All Classes		0.696	0.652	0.674	0.673	0.784	0.658	0.720	0.703	0.738	0.725	0.736	0.727
Coats	Correct utilization	0.666	0.837	0.792	0.742	0.743	0.809	0.819	0.774	0.689	0.840	0.795	0.757
	Incorrect utilization	0.759	0.780	0.817	0.769	0.819	0.784	0.846	0.801	0.841	0.838	0.868	0.840
Earmuffs	Correct utilization	0.485	0.474	0.426	0.479	0.539	0.400	0.480	0.459	0.565	0.411	0.484	0.476
	Incorrect utilization	0.685	0.551	0.606	0.611	0.780	0.663	0.711	0.698	0.682	0.694	0.650	0.688
Gloves	Correct utilization	0.723	0.685	0.707	0.703	0.783	0.667	0.752	0.720	0.760	0.746	0.769	0.753
	Incorrect utilization	0.610	0.420	0.452	0.497	0.788	0.379	0.494	0.512	0.606	0.583	0.555	0.595
Hard hats	Correct utilization	0.730	0.899	0.870	0.806	0.818	0.839	0.881	0.829	0.794	0.876	0.881	0.833
	Incorrect utilization	0.602	0.717	0.744	0.654	0.746	0.736	0.803	0.741	0.646	0.792	0.761	0.712
Glasses	Correct utilization	0.763	0.730	0.764	0.746	0.807	0.716	0.799	0.759	0.826	0.780	0.824	0.802
	Incorrect utilization	0.727	0.729	0.783	0.728	0.850	0.746	0.787	0.795	0.798	0.792	0.814	0.795
Masks	Correct utilization	0.826	0.864	0.899	0.845	0.886	0.876	0.921	0.881	0.869	0.893	0.928	0.881
	Incorrect utilization	0.781	0.796	0.822	0.788	0.853	0.815	0.847	0.834	0.857	0.852	0.874	0.855
Safety shoes	Correct utilization	0.809	0.148	0.239	0.250	0.799	0.259	0.320	0.392	0.659	0.407	0.428	0.504
	Incorrect utilization	0.581	0.491	0.512	0.532	0.765	0.554	0.619	0.642	0.738	0.646	0.672	0.689

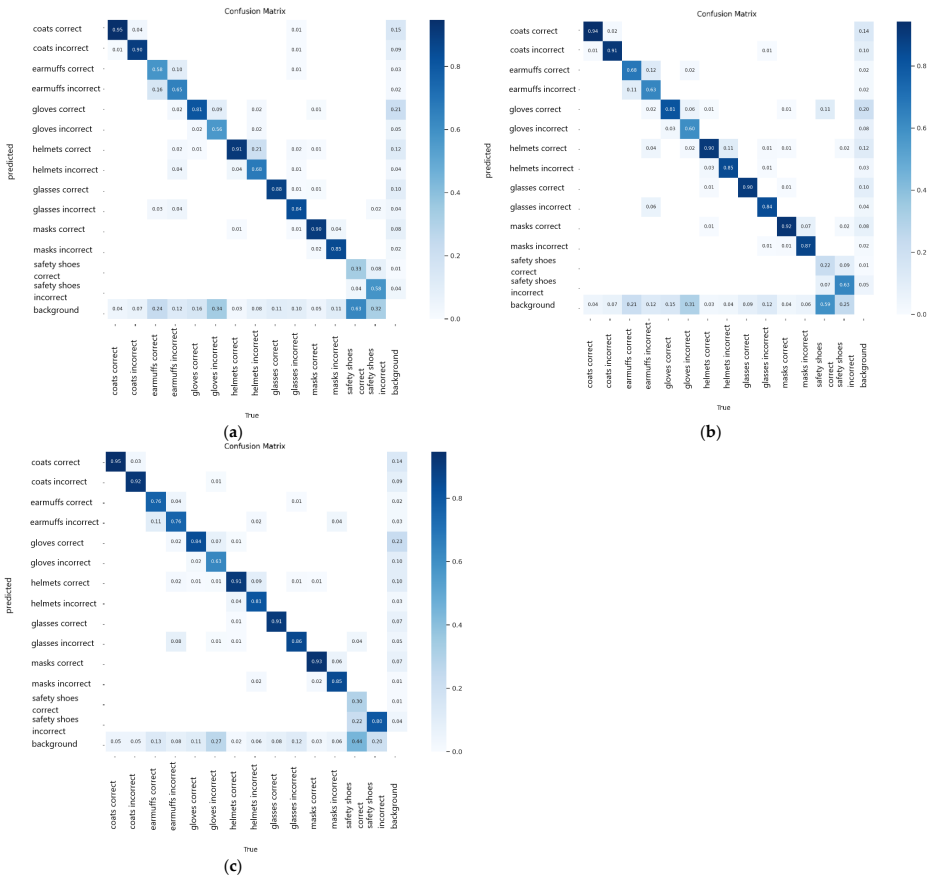


Figure 20. Confusion matrix of YOLOv6 models. (**a**) 50 epochs model, (**b**) 75 epochs model, and (**c**) 100 epochs model.

Similar to the other models, there is low performance of the YOLOv6 models in detecting certain PPE classes, which are earmuffs correct, hard hats incorrect, and safety shoes correct utilization. The low metric values for those classes are mainly attributed to an

imbalance of the number of training images. Nevertheless, the overall performance of the YOLOv6 models is significantly better than the YOLOv4 models.

Based on the results of the detection model testing using YOLOv6, the best performance was achieved with the YOLOv6 model with the number of epochs set at 100. This is because it yielded the highest values for the evaluation metrics, including recall, mAP, F1 score, and accuracy, when compared to the 50 epochs and 75 epochs models. While the precision value of the 100 epochs model was lower than those of the 75 epochs model, the model with the highest F1 score is considered the best choice because it represents a balance between precision and recall. Therefore, epoch 100 is chosen as the optimal epoch.

The precision result for the best model, which is the 100 epochs model, is 0.738, meaning that approximately 73.8% of objects identified as positive by the model are indeed relevant objects. The recall result is 0.725, indicating that the model can detect approximately 72.5% of all true positive objects. The F1 score, which combines precision and recall to provide an overall measure of model performance, is 0.727, showing that the model maintains a good balance between precision and recall.

The mAP is a value that combines multiple precision–recall values at different decision thresholds into a single average score. The mAP result is 0.736, indicating that the model performs well in detection across various decision thresholds. Accuracy represents the overall percentage of correct predictions out of all predictions. In the YOLOv6 detection model with 100 epochs, the accuracy achieved is 82.02%, indicating that the model correctly predicts approximately 82.02% of all test dataset examples. In conclusion, the YOLOv6 detection model with the number of epochs set to 100 demonstrates strong performance with adequate values for precision, recall, mAP, F1 score, and accuracy.

5.4. Comparison of the Models

Based on the results obtained from the model of each algorithm (the 100 epochs models of YOLOv4, YOLOv5, and YOLOv6), a comparison was conducted to determine the best PPE detection model. The criteria for selecting the best detection model are based on precision, recall, mean Average Precision (mAP), and F1 score values. The choice of evaluation metrics is crucial because these values consider the model's accuracy and its ability to correctly identify various object classes. Accuracy is not considered a determining factor for assessing model success since it can be biased and may not provide an accurate representation of object detection quality when there is an imbalance in the number of objects in each class. In this case, the number of objects in each class is not uniform and balanced. The performance comparison of the best model using each algorithm on the validation dataset is presented in Table 8.

Table 8. Performance comparison of the best model using each algorithm.

Model	Precision	Recall	mAP	F1 Score
YOLOv4 100 epochs	0.643	0.624	0.657	0.633
YOLOv5 100 epochs	0.739	0.750	0.757	0.744
YOLOv6 100 epochs	0.738	0.725	0.736	0.727

Therefore, when examining precision, recall, mAP, and F1 score values, it is evident that the best model is YOLOv5, scoring the best values in each evaluation metric. However, the differences in values between YOLOv5 and YOLOv6 in each metric are not significantly large, so both algorithms can be considered equally effective in object detection. Meanwhile, YOLOv4 demonstrates performance below the other two algorithms, indicating that the YOLOv4 model is still not quite effective in detecting PPE objects.

In addition to having the highest metric values for PPE object detection, YOLOv5 also boasts faster computation times for both training and testing compared to the other two algorithms. The testing time of each model is presented in Table 9. The average testing time is measured for detecting PPE objects in a single image. The time taken during testing indicates the model's ability to perform detection at varying speeds. When a model can

detect objects quickly, especially when measured in milliseconds, it can be considered to have real-time detection capabilities. In this research, the fastest time for detection recorded was 1.02 s, indicating that the model is quite proficient at detection when considering its speed in recognizing objects. Noted that testing time presented in the table is not only the time taken for making a prediction, but also includes the time for importing the data and preprocessing the image.

Table 9. Average testing time of all detection models.

Algorithms	Number of Epochs	Testing Time per Image
YOLOv4	50	1.80 s
	75	1.45 s
	100	1.40 s
YOLOv5	50	1.15 s
	75	1.02 s
	100	1.03 s
YOLOv6	50	1.32 s
	75	1.11 s
	100	1.05 s

5.5. Discussion

The object detection system is designed to monitor the use of PPE in real-time and issue alerts for any incorrect usage or absence of PPE. These cameras are strategically positioned at various angles to encompass the front, rear, and side areas surrounding the machinery within manufacturing teaching laboratories in universities. To ensure comprehensive coverage within the detection zone, a single camera is capable of effectively detecting objects within an area of approximately 15 m^2. Given this coverage capacity, the camera can encompass up to three machines within its field of view.

The detection of PPE using YOLOv4, YOLOv5, and YOLOv6 has yielded results that indicate YOLOv5 with 100 epochs as the best detection model when compared to the other two YOLO variants. This conclusion is based on the evaluation metric values, which show that the YOLOv5 model outperforms the other models across all four evaluation metrics, despite having slightly lower accuracy than YOLOv6 with 100 epochs. Examples of the results of the best model for each YOLO version are presented in Figures 21 and 22. It is observed that there are some errors and instances of missed detections in the models. In the single individual image, the YOLOv4 and YOLOv5 models are able to find all PPE objects and detect them as the correct classes. Meanwhile, the YOLOv6 model fails to detect earmuff objects.

The situations that cause errors when detecting are when an object is covered by another object and when a PPE object is captured by a small camera. To overcome this, cameras can be placed on various sides that include the machine and its users. As such, the usage of PPE can be constantly monitored. When detecting a PPE object that has a similarity to the background, the model encountered a detection error. In order to overcome this, there is a need to augment training datasets with various backgrounds or more complex location backgrounds to avoid the inability to detect objects of varying colors and shapes.

Another error is that when there is a PPE object that is not used by humans but is present in the camera capture (e.g., PPE is on the table), then the detection model will detect that there is a PPE object. This is an error when implemented as a warning signal for PPE usage alerts. To resolve this problem, adding labels or annotations is required for PPE objects that are not used by humans.

In the images with multiple individuals, YOLOv5 could not detect some of the gloves, masks, and glasses, and the model mistakenly classifies machine parts as safety shoes incorrect. In YOLOv6, errors include misclassifying machine parts as safety shoes incorrect, and missed detections of glasses, masks, and gloves. However, in the case of YOLOv6, the

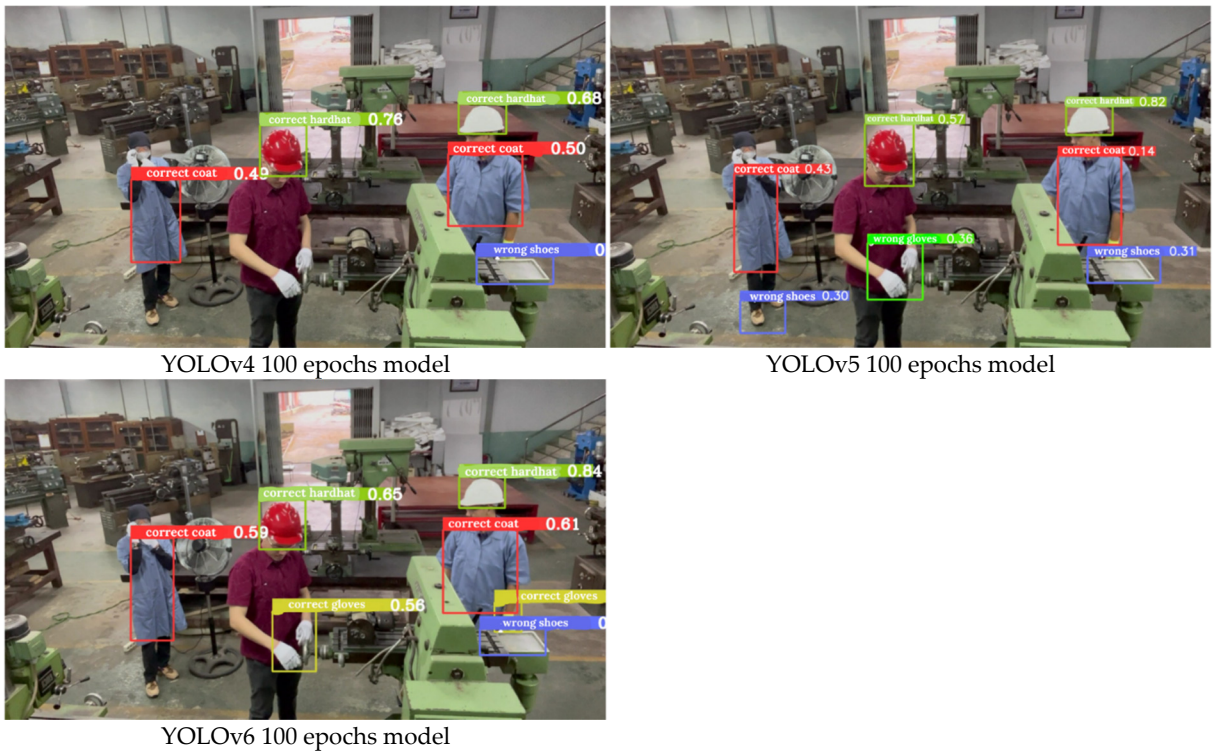
detection results for the multiple individual image show that gloves are correctly detected on the third individual (on the far right), whereas YOLOv5 fails to detect them. Errors in detection, such as machine parts being classified as safety shoes incorrect, occur due to the presence of objects that resemble the background, affecting the detection capability. This is because visual-based detection relies on similarities in color and shape, which can lead to misclassifications. Another factor is the imbalance in the number of objects from different classes in the training data, causing the model to be biased towards the class with the higher number of objects. This can result in higher detection errors for classes with fewer objects, such as the safety shoes correct objects which are frequently mistaken as incorrect. Additionally, during the annotation process, if there are objects outside the designated class that fall within the bounding box, it can confuse the model and lead to incorrect detections of background objects that resemble the target class.

In this study, all proposed CNN models have data limitations, meaning they can only detect PPE objects that were part of the training dataset (the 14 PPE classes). After conducting further experiments, it is found that the models cannot detect other types of PPE objects because the deep learning models built in this research are based on supervised learning, where the model can only detect objects based on what it learned from the training datasets. To enhance the detection capabilities of the deep learning models, it is advisable to introduce data variations or different types of PPE objects into the training and testing datasets for further research, not only in the setting of manufacturing teaching laboratories, but also other science (e.g., chemical and biological) laboratories.

This study also found that class imbalance can affect the performance of the detection model, especially in underrepresented classes. In addressing class imbalances, data resampling can be performed. The technique used is to oversample the number of samples in the minority class by retrieving new data in the minority class and replicating or making copies of existing data. This would enhance the deep learning model's ability to learn more efficiently, thereby boosting its performance in detecting objects.

YOLOv4 100 epochs model YOLOv5 100 epochs model YOLOv6 100 epochs model

Figure 21. Comparison of detection results on single individual sample.



YOLOv4 100 epochs model

YOLOv5 100 epochs model

YOLOv6 100 epochs model

Figure 22. Comparison of detection results on sample with multiple individuals.

Adapting this system for real-world university settings, it could be integrated with monitoring cameras within laboratory environments, akin to a CCTV setup equipped with an alert system—for instance, issuing a warning notice when PPE is not being used or is misused by individuals. This real-time monitoring ensures continuous detection and intervention, making it a practical and effective safety tool. The system only necessitates investment in surveillance cameras and computer devices, thus it is relatively easier to be implemented in the laboratory. Its primary advantage lies in minimizing both physical harm and financial losses related to accidents and medical care for university affiliates, thereby enhancing safety and reducing potential liabilities.

On the other side of the spectrum, enhancing the education and training of laboratory staff, students, and visitors remains a pivotal strategy for improving the correct utilization of PPE within manufacturing teaching laboratories. This approach not only involves instructing users on the correct selection and fitting of PPE but also extends to educating them about its limitations and the necessity for regular inspection to ensure its integrity and effectiveness. Comprehensive and efficient training programs, tailored to address the specific hazards encountered in these environments, play a crucial role in fostering a deeper understanding of safety protocols and the underlying reasons for their implementation.

6. Conclusions

This research presents a visual-based detection approach for images to identify the completeness and correctness of PPE usage in manufacturing teaching laboratories in universities. The created detection model aims not only to detect the PPE objects being used but also to assess their proper usage. There are seven types of PPE objects examined in this study, each with two conditions: proper usage and improper usage, resulting in a total of 14 object classes. Three algorithms were proposed for this study: YOLOv4, YOLOv5, and YOLOv6, to build the detection models. For each algorithm, three models were constructed, with 50, 75, and 100 epochs.

The performance of the proposed approach was evaluated based on the choice of algorithm used to obtain the best-performing detection model. The experimental results

demonstrate that all proposed deep learning models can effectively detect various classes of PPE to assess the completeness and correctness of their usage in manufacturing teaching laboratories. YOLOv5 with 100 epochs exhibited superior performance compared to YOLOv4 and YOLOv6, as determined by the comparison of evaluation metric values obtained from testing with the dataset.

Some suggestions for potential future enhancements in object detection systems include the following: (1) Expanding the training dataset, particularly for shoes and earmuffs objects, to provide a larger volume of data. This would enable the deep learning model to learn more effectively and improve its detection performance. (2) Increasing the dataset's diversity by including variations in color, shape, and material for each PPE object. This would empower the model to better detect a wider range of PPE variations.

Author Contributions: Conceptualization, A.P.R.; methodology, A.S.L. and A.P.R.; software, A.S.L.; validation, A.S.L. and A.P.R.; formal analysis, A.S.L. and A.P.R.; investigation, A.S.L. and A.P.R.; resources, A.P.R.; data curation, A.P.R.; writing—original draft preparation, A.S.L. and A.P.R.; writing—review and editing, A.S.L. and A.P.R.; visualization, A.S.L. and A.P.R.; supervision, A.P.R.; project administration, A.P.R.; funding acquisition, A.P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Universitas Gadjah Mada Final Project Recognition Program Grant for Fiscal Year 2023 (RTA 2023) based on assignment letter number 5075/UN1.P.II/DitLit/PT.01.01/2023.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Written informed consent has been obtained from the respondents to publish this paper.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We would like to express our sincere thanks to the manufacturing laboratory staff and all respondents who provide the objects and assist during the data collection.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

References

1. BPJS Ketenagakerjaan. Laporan Keuangan dan Pengelolaan Program BPJS Ketenagakerjaan 2021. 2022. Available online: https://www.bpjsketenagakerjaan.go.id/assets/uploads/laporan_keuangan/LK_LPP_BPJAMSOSTEK_2021.pdf (accessed on 12 February 2023).
2. Vasconcelos, B.; Junior, B.B. The Causes of Workplace Accidents and their Relation to Construction Equipment Design. *Procedia Manuf.* **2015**, *3*, 4392–4399. [CrossRef]
3. Alamneh, Y.M.; Wondifraw, A.Z.; Negesse, A.; Ketema, D.B.; Akalu, T.Y. The prevalence of occupational injury and its associated factors in Ethiopia: A systematic review and meta-analysis. *J. Occup. Med. Toxicol.* **2020**, *15*, 14. [CrossRef]
4. Baye, B.F.; Baye, M.F.; Teym, A.; Derseh, B.T. Utilization of Personal Protective Equipment and Its Associated Factors among Large Scale Factory Workers in Debre Berhan Town, Ethiopia. *Environ. Health Insights* **2022**, *16*, 1–9. [CrossRef]
5. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [CrossRef] [PubMed]
6. Nath, N.D.; Behzadan, A.H.; Paal, S.G. Deep learning for site safety: Real-time detection of personal protective equipment. *Autom. Constr.* **2020**, *112*, 103085. [CrossRef]
7. Akbar-Khanzadeh, F. Factors contributing to discomfort or dissatisfaction as a result of wearing personal protective equipment. *J. Hum. Ergol.* **1998**, *27*, 70–75. [CrossRef]
8. Xiong, R.; Tang, P. Pose guided anchoring for detecting proper use of personal protective equipment. *Autom. Constr.* **2021**, *130*, 103828. [CrossRef]
9. Vukicevic, A.M.; Djapan, M.; Isailovic, V.; Milasinovic, D.; Savkovic, M.; Milosevic, P. Generic compliance of industrial PPE by using deep learning techniques. *Saf. Sci.* **2022**, *148*, 105646. [CrossRef]
10. Buchweiller, J.P.; Mayer, A.; Klein, R.; Iotti, J.M.; Kusy, A.; Reinert, D.; Christ, E. Safety of electronic circuits integrated into personal protective equipment (PPE). *Saf. Sci.* **2003**, *41*, 395–408. [CrossRef]

11. Cheng, J.C.P.; Wong, P.K.Y.; Luo, H.; Wang, M.; Leung, P.H. Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Autom. Constr.* **2022**, *139*, 104312. [CrossRef]
12. Saudi, M.M.; Ma'arof, A.H.; Ahmad, A.; Saudi, A.S.M.; Ali, M.H.; Narzullaev, A.; Ghazali, M.I.M. Image detection model for construction worker safety conditions using faster R-CNN. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 246–250. [CrossRef]
13. Lo, J.H.; Lin, L.K.; Hung, C.C. Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm. *Sustainability* **2023**, *15*, 391. [CrossRef]
14. Ngoc-Thoan, N.; Bui, D.Q.T.; Tran, C.N.N.; Tran, D.H. Improved detection network model based on YOLOv5 for warning safety in construction sites. *Int. J. Constr. Manag.* **2023**, *1*, 11. [CrossRef]
15. Gallo, G.; Di Rienzo, F.; Ducange, P.; Ferrari, V.; Tognetti, A.; Vallati, C. A Smart System for Personal Protective Equipment Detection in Industrial Environments Based on Deep Learning. In Proceedings of the 2021 IEEE International Conference on Smart Computing, SMARTCOMP, Irvine, CA, USA, 23–27 August 2021; pp. 222–227. [CrossRef]
16. Cheng, R. A survey: Comparison between Convolutional Neural Network and YOLO in image identification. *J. Phys. Conf. Ser.* **2020**, *1453*, 012139. [CrossRef]
17. Ji, X.; Gong, F.; Yuan, X.; Wang, N. A high-performance framework for personal protective equipment detection on the offshore drilling platform. *Complex Intell. Syst.* **2023**, *9*, 5637–5652. [CrossRef]
18. Protik, A.A.; Rafi, A.H.; Siddique, S. Real-time Personal Protective Equipment (PPE) Detection Using YOLOv4 and TensorFlow. In Proceedings of the TENSYMP 2021—2021 IEEE Region 10 Symposium, Jeju, Republic of Korea, 23–25 August 2021. [CrossRef]
19. Collo, M.L.R.; Richard, M.; Esguerra, J.; Sevilla, R.V.; Malunao, D.C. A COVID-19 Safety Monitoring System: Personal Protective Equipment (PPE) Detection using Deep Learning. In Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications, DASA, Chiangrai, Thailand, 23–25 March 2022; pp. 295–299. [CrossRef]
20. Yung, N.D.T.; Wong, W.K.; Juwono, F.H.; Sim, Z.A. Safety Helmet Detection Using Deep Learning: Implementation and Comparative Study Using YOLOv5, YOLOv6, and YOLOv7. In Proceedings of the 2022 International Conference on Green Energy, Computing and Sustainable Technology, GECOST, Miri Sarawak, Malaysia, 26–28 October 2022; pp. 164–170. [CrossRef]
21. Schröder, I.; Huang, D.Y.Q.; Ellis, O.; Gibson, J.H.; Wayne, N.L. Laboratory safety attitudes and practices: A comparison of academic, government, and industry researchers. *J. Chem. Health Saf.* **2016**, *23*, 12–23. [CrossRef]
22. Gopalaswami, N.; Han, Z. Analysis of laboratory incident database. *J. Loss Prev. Process Ind.* **2020**, *64*, 104027. [CrossRef]
23. Kumar, S.; Gupta, H.; Yadav, D.; Ansari, I.A.; Verma, O.P. YOLOv4 algorithm for the real-time detection of fire and personal protective equipments at construction sites. *Multimed. Tools Appl.* **2022**, *81*, 22163–22183. [CrossRef]
24. Kwak, N.J.; Kim, D.J. Detection of Worker's Safety Helmet and Mask and Identification of Worker Using Deeplearning. *Comput. Mater. Contin.* **2023**, *75*, 1671–1686. [CrossRef]
25. Norkobil Saydirasulovich, S.; Abdusalomov, A.; Jamil, M.K.; Nasimov, R.; Kozhamzharova, D.; Cho, Y.I. A YOLOv6-Based Improved Fire Detection Approach for Smart City Environments. *Sensors* **2023**, *23*, 3161. [CrossRef]
26. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
27. Diwan, T.; Anirudh, G.; Tembhurne, J.V. Object detection using YOLO: Challenges, architectural successors, datasets and applications. *Multimed. Tools Appl.* **2023**, *82*, 9243–9275. [CrossRef] [PubMed]
28. Wang, Z.; Wu, Y.; Yang, L.; Thirunavukarasu, A.; Evison, C.; Zhao, Y. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors* **2021**, *21*, 3478. [CrossRef] [PubMed]
29. Li, J.; Zhao, X.; Zhou, G.; Zhang, M. Standardized use inspection of workers' personal protective equipment based on deep learning. *Saf. Sci.* **2022**, *150*, 105689. [CrossRef]
30. Kasper-Eulaers, M.; Hahn, N.; Kummervold, P.E.; Berger, S.; Sebulonsen, T.; Myrland, Ø. Short Communication: Detecting Heavy Goods Vehicles in Rest Areas in Winter Conditions Using YOLOv5. *Algorithms* **2021**, *14*, 114. [CrossRef]
31. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [CrossRef]